# Technical Report # KU-EC-09-3: A Probabilistic Characterization of Random Proximity Catch Digraphs and the Associated Tools

Elvan Ceyhan *

March 28, 2009

**Abstract**

Proximity catch digraphs (PCDs) are based on proximity maps which yield proximity regions and are special types of proximity graphs. PCDs are based on the relative allocation of points from two or more classes in a region of interest and have applications in various fields. In this article, we provide auxiliary tools for and various characterizations of PCDs based on their probabilistic behavior. We consider the cases in which the vertices of the PCDs come from uniform and non-uniform distributions in the region of interest. We also provide some of the newly defined proximity maps as illustrative examples.

*Keywords:* class cover catch digraph (CCCD); central similarity PCD; Delaunay triangulation; domination number; proportional-edge PCD; proximity graph; random graph; relative arc density

*Address: Department of Mathematics, Koç University, 34450 Sarıyer, Istanbul, Turkey. e-mail: elceyhan@ku.edu.tr, tel:+90 (212) 338-1845, fax: +90 (212) 338-1559.

# 1 Introduction

The proximity catch digraphs (PCDs) are a special type of proximity graphs which are based on proximity maps and are used in disciplines where shape and structure are crucial. Examples include computer vision (dot patterns), image analysis, pattern recognition (prototype selection), geography and cartography, visual perception, biology, etc. *Proximity graphs* were first introduced by Toussaint (1980), who called them *relative neighborhood graphs*. The notion of relative neighborhood graph has been generalized in several directions and all of these graphs are now called proximity graphs. From a mathematical and algorithmic point of view, proximity graphs fall under the category of *computational geometry*.

In recent years, a new classification and spatial pattern analysis approach which is based on the relative positions of the data points from various classes has been developed. Priebe et al. (2001) introduced the class cover catch digraphs (CCCDs) and gave the exact and the asymptotic distribution of the domination number of the CCCD based on two data sets $\mathcal{X}_n$ and $\mathcal{Y}_m$ both of which are random samples from uniform distribution on a compact interval in $\mathbb{R}$. DeVinney et al. (2002), Marchette and Priebe (2003), Priebe et al. (2003b), Priebe et al. (2003a), and DeVinney and Priebe (2006) applied the concept in higher dimensions and demonstrated relatively good performance of CCCD in classification. The employed methods involve data reduction (condensing) by using approximate minimum dominating sets as prototype sets, since finding the exact minimum dominating set is in general an NP-hard problem — in particular, for CCCD — (see DeVinney (2003)). Furthermore, the exact and the asymptotic distribution of the domination number of the CCCDs are not analytically tractable in higher dimensions. Ceyhan (2004) extended the concept of CCCDs by introducing PCDs, which do not suffer from some of the shortcomings of CCCDs in higher dimensions. In particular, two new types of PCDs (namely, *proportional-edge* and *central similarity PCDs*) are introduced; distribution of the domination number of proportional-edge PCDs is calculated, and is applied in testing spatial patterns of segregation and association (Ceyhan and Priebe (2005, 2007)). The distributions of the relative arc density of these PCD families are also derived and used for the same purpose (Ceyhan et al. (2006) and Ceyhan et al. (2007)).

A general definition of proximity graphs is as follows: Let $V$ be any finite or infinite set of points in $\mathbb{R}^d$. Each (unordered) pair of points $(p, q) \in V \times V$ is associated with a neighborhood $\mathfrak{N}(p, q) \subseteq \mathbb{R}^d$. Let $\mathfrak{P}$ be a property defined on $\mathfrak{N} = \{\mathfrak{N}(p, q) : (p, q) \in V \times V\}$. A *proximity* (or *neighborhood*) *graph* $G_{\mathfrak{N},\mathfrak{P}}(V, E)$ defined by the property $\mathfrak{P}$ is a graph with the set of vertices $V$ and the set of edges $E$ such that $(p, q) \in E$ iff $\mathfrak{N}(p, q)$ satisfies property $\mathfrak{P}$. Examples of most commonly used proximity graphs are the Delaunay tessellation, the boundary of the convex hull, the Gabriel graph, relative neighborhood graph, Euclidean minimum spanning tree, and sphere of influence graph of a finite data set. See, e.g., Jaromczyk and Toussaint (1992).

The relative allocation of the data points are used to construct a proximity digraph. A *digraph* is a directed graph, i.e., a graph with directed edges from one vertex to another based on a binary relation. Then the pair $(p, q) \in V \times V$ is an ordered pair and $(p, q)$ is an arc (directed edge) denoted $pq$ to reflect its difference from an edge. For example, the nearest neighbor (di)graph in Paterson and Yao (1992) is a proximity digraph. The nearest neighbor digraph, denoted $NND(V)$, has the vertex set $V$ and $pq$ an arc iff $d(p, q) = \min_{v \in V \setminus \{p\}} d(p, v)$. That is, $pq$ is an arc of $NND(V)$ iff $q$ is a nearest neighbor of $p$. Note that if $pq$ is an arc in $NND(V)$, then $(p, q)$ is an edge in $RNG(V)$. Our PCDs are based on the property $\mathfrak{P}$ that is determined by the following mapping which is defined in a more general space than $\mathbb{R}^d$. Let $(\Omega, \mathcal{M})$ be a measurable space. The *proximity map* $N(\cdot)$ is given by $N : \Omega \to \wp(\Omega)$, where $\wp(\cdot)$ is the power set functional, and the *proximity region* of $x \in \Omega$, denoted $N(x)$, is the image of $x \in \Omega$ under $N(\cdot)$. The points in $N(x)$ are thought of as being "closer" to $x \in \Omega$ than are the points in $\Omega \setminus N(x)$. Proximity maps are the building blocks of the *proximity graphs* of Toussaint (1980); an extensive survey is available in Jaromczyk and Toussaint (1992).

The *PCD* $D = (\mathcal{V}, \mathcal{A})$ has the vertex set $\mathcal{V} = \{p_1, p_2, \ldots, p_n\}$ and the arc set $\mathcal{A}$ is defined by $p_i p_j \in \mathcal{A}$ iff $p_j \in N(p_i)$ for $i \neq j$. Notice that $D$ depends on the *proximity* map $N(\cdot)$, and if $p_j \in N(p_i)$, then $N(p_i)$ is said to *catch* $p_j$. Hence the name *proximity catch digraph*. If arcs of the form $p_i p_i$ (i.e., loops) were allowed, $D$ would have been called a *pseudodigraph* according to some authors (see, e.g., Chartrand and Lesniak

(1996)).

In this article, we provide a probabilistic characterization of the proximity maps, and the associated regions and PCDs, and introduce auxiliary tools for the PCDs. We define the proximity maps and data-random PCDs in Section 2, describe the auxiliary tools (such as edge and vertex regions) for the construction of PCDs in Section 3, provide $\Gamma_1$-regions and the related concepts for proximity maps in Section 4, discuss the examples of proximity maps in Delaunay triangles in Section 6, and provide the transformations preserving uniformity on triangles in $\mathbb{R}^2$ in Section 5. We investigate the characterization of proximity regions and the associated PCDs in Section 7, introduce $\Gamma_k$-regions for proximity maps in $T(\mathcal{Y}_3)$ in Section 8, $\kappa$-values for the proximity maps in $T(\mathcal{Y}_3)$ in Section 9, and provide discussion and conclusions in Section 10.

## 2 Proximity Maps and Data-Random PCDs

Let $\mathcal{X}_n = \{X_1, X_2, \ldots, X_n\}$ and $\mathcal{Y}_m = \{Y_1, Y_2, \ldots, Y_m\}$ be two data sets from classes $\mathcal{X}$ and $\mathcal{Y}$ of $\Omega$-valued random variables whose joint pdf is $f_{X,Y}$. Let $d(\cdot,\cdot) : \Omega \times \Omega \to [0,\infty)$ be a distance function. The class cover problem for a target class, say $\mathcal{X}_n$, refers to finding a collection of neighborhoods, $N(X_i)$ around $X_i \in \mathcal{X}_n$ such that (i) $\mathcal{X}_n \subset \left(\bigcup_i N(X_i)\right)$ and (ii) $\mathcal{Y}_m \cap \left(\bigcup_i N(X_i)\right) = \emptyset$. A collection of neighborhoods satisfying both conditions is called a *class cover*. A cover satisfying condition (i) is a *proper cover* of class $\mathcal{X}$ while a collection satisfying condition (ii) is a *pure cover* relative to class $\mathcal{Y}$. From a practical point of view, for example for classification, of particular interest are the class covers satisfying both (i) and (ii) with the smallest collection of neighborhoods, i.e., minimum cardinality cover. This class cover problem is a generalization of the set cover problem in Garfinkel and Nemhauser (1972) that emerged in statistical pattern recognition and machine learning, where an edited or condensed set (prototype set) is selected from $\mathcal{X}_n$ (see, e.g., Devroye et al. (1996)).

In particular, we construct the proximity regions using data sets from two classes. Given $\mathcal{Y}_m \subseteq \Omega$, the *proximity map* $N_{\mathcal{Y}}(\cdot)$ associates a *proximity region* $N_{\mathcal{Y}}(x) \subseteq \Omega$ with each point $x \in \Omega$. The region $N_{\mathcal{Y}}(x)$ is defined in terms of the distance between $x$ and $\mathcal{Y}_m$. More specifically, our proximity maps will be based on the relative position of points from class $\mathcal{X}$ with respect to the Delaunay tessellation of the class $\mathcal{Y}$. See Okabe et al. (2000) and Ceyhan (2009) for more on Delaunay tessellations.

If $\mathcal{X}_n$ is a set of $\Omega$-valued random variables then $N_{\mathcal{Y}}(X_i)$ are random sets. If $X_i$ are independent identically distributed then so are the random sets $N_{\mathcal{Y}}(X_i)$. We define the data-random PCD $D = (\mathcal{V}, \mathcal{A})$ — associated with $N_{\mathcal{Y}}(\cdot)$ — with vertex set $\mathcal{V} = \mathcal{X}_n$ and arc set $\mathcal{A}$ by $X_i X_j \in \mathcal{A} \iff X_j \in N_{\mathcal{Y}}(X_i)$. Since this relationship is not symmetric, a digraph is needed rather than a graph. The random digraph $D$ depends on the (joint) distribution of the $X_i$ and on the map $N_{\mathcal{Y}}(\cdot)$.

The PCDs are closely related to the *proximity graphs* of Jaromczyk and Toussaint (1992) and might be considered as a special case of *covering sets* of Tuza (1994) and *intersection digraphs* of Sen et al. (1989). This data random proximity digraph is a *vertex-random proximity digraph* which is not of standard type. The randomness of the PCDs lies in the fact that the vertices are random with joint pdf $f_{X,Y}$, but arcs $X_i X_j$ are deterministic functions of the random variable $X_j$ and the set $N_{\mathcal{Y}}(X_i)$.

For example, the CCCD of Priebe et al. (2001) can be viewed as an example of PCD with $N_{\mathcal{Y}}(x) = B(x, r(x))$, where $r(x) := \min_{\mathsf{y} \in \mathcal{Y}_m} d(x, \mathsf{y})$. The CCCD is the digraph of order $n$ with vertex set $\mathcal{X}_n$ and an arc from $X_i$ to $X_j$ iff $X_j \in B(X_i, r(X_i))$. That is, there is an arc from $X_i$ to $X_j$ iff there exists an open ball centered at $X_i$ which is "pure" (or contains no elements) of $\mathcal{Y}_m$, and simultaneously contains (or "catches") point $X_j$.

## 3 Auxiliary Tools for the Construction of PCDs in $\mathbb{R}^d$

Recall the proximity map (associated with CCCD) in $\mathbb{R}$ is defined as $B(x, r(x))$ where $r(x) = \min_{y \in \mathcal{Y}_m} d(x, y)$ with $d(x, y)$ being the Euclidean distance between $x$ and $y$ (Priebe et al. (2001)). Our goal is to extend this

idea to higher dimensions and investigate the associated digraph. Now let $\mathcal{Y}_m = \{\mathsf{y}_1, \mathsf{y}_2, \dots, \mathsf{y}_m\} \subset \mathbb{R}^d$. For $d = 1$ the proximity map associated with CCCD is defined as the open ball $N_S(x) := B(x, r(x))$ for all $x \in \mathbb{R} \setminus \mathcal{Y}_m$ and for $x \in \mathcal{Y}_m$, define $N_S(x) = \{x\}$. Furthermore, dependence on $\mathcal{Y}_m$ is through $r(x)$. Hence $N_S(x)$ is based on the intervals $I_{i-1} = \left(\mathsf{y}_{(i-1):m}, \mathsf{y}_{i:m}\right)$ for $i = 1, 2, \dots, (m+1)$ with $\mathsf{y}_{0:m} = -\infty$ and $\mathsf{y}_{(m+1):m} = \infty$ where $\mathsf{y}_{i:m}$ is the $i^{th}$ order statistic in $\mathcal{Y}_m$. This intervalization can be viewed as a tessellation since it partitions $C_H(\mathcal{Y}_m)$, the convex hull of $\mathcal{Y}_m$. For $d > 1$, a natural tessellation that partitions $C_H(\mathcal{Y}_m)$ is the Delaunay tessellation (see Okabe et al. (2000) and Ceyhan (2009)). Let $\mathcal{T}_i$ for $i = 1, 2, \dots, J$ be the $i^{th}$ Delaunay cell in the Delaunay tessellation of $\mathcal{Y}_m$. In $\mathbb{R}$, we implicitly use the cell that contains $x$ to define the proximity map.

A natural extension of the proximity region $N_S(x)$ to multiple dimensions (i.e., to $\mathbb{R}^d$ with $d > 1$) is obtained by the same definition as above; that is, $N_S(x) := B(x, r(x))$ where $r(x) := \min_{\mathsf{y} \in \mathcal{Y}_m} d(x, \mathsf{y})$. Notice that a ball is a sphere in higher dimensions, hence the name *spherical proximity map* and the notation $N_S$. The spherical proximity map $N_S(x)$ is well-defined for all $x \in \mathbb{R}^d$ provided that $\mathcal{Y}_m \neq \emptyset$. Extensions to $\mathbb{R}^2$ and higher dimensions with the spherical proximity map — with applications in classification — are investigated in DeVinney et al. (2002), DeVinney and Wierman (2003), Marchette and Priebe (2003), Priebe et al. (2003a), Priebe et al. (2003b), and DeVinney and Priebe (2006). However, finding the minimum dominating set of the PCD associated with $N_S(\cdot)$ is an NP-hard problem and the distribution of the domination number is not analytically tractable for $d > 1$ (Ceyhan (2004)). This drawback has motivated us to define new types of proximity maps in higher dimensions. Note that for $d = 1$, such problems do not occur. Ceyhan (2009) states some appealing properties of the proximity map $N_S(x) = B(x, r(x))$ in $\mathbb{R}$ and uses them as guidelines for extending proximity maps to higher dimensions and defining new proximity maps. After a slight modification, the spherical proximity maps gives rise to arc-slice proximity maps which is defined as $N_{AS}(x) := B(x, r(x)) \cap T(\mathcal{Y}_3)$ for $x \in T(\mathcal{Y}_3)$ (i.e., when the arc-slice proximity region is the spherical proximity region restricted to the Delaunay triangle $x$ lies in). However, for $x \notin C_H(\mathcal{Y}_m)$ (i.e., $x$ is not in any of the Delaunay triangles based on $\mathcal{Y}_m$), $N_{AS}(x)$ is not defined.

For $x \in I_i$, $N_S(x) = I_i$ iff $x = \left(\mathsf{y}_{(i-1):m} + \mathsf{y}_{i:m}\right)/2$. We define an associated region for such points in the general context.

**Definition 3.1.** The *superset region* for any proximity map $N(\cdot)$ in $\Omega$ is defined to be $\mathscr{R}_S(N) := \left\{x \in \Omega : N(x) = \Omega\right\}$. When $X$ is $\Omega$-valued random variable, then we assume $X \in \mathscr{R}_S(N)$ if $N(X) = \Omega$ a.s. $\square$

For example, for $\Omega = I_i \subsetneq \mathbb{R}$ with $i = 1, 2, \dots, (m-1)$, $\mathscr{R}_S(N_S) := \{x \in I_i : N_S(x) = I_i\} = \left\{\left(\mathsf{y}_{(i-1):m} + \mathsf{y}_{i:m}\right)/2\right\}$, and for $i = 0, m$ (i.e., $\Omega = I_0$ or $\Omega = I_m$, then $\mathscr{R}_S(N_S) = \emptyset$ since $N_S(x) \subsetneq I_i$ for all $x \in I_i$ for $i = 0, m$. More generally for $\Omega = \mathcal{T}_i \subsetneq \mathbb{R}^d$ (i.e., $i^{th}$ Delaunay cell), $\mathscr{R}_S(N_S) := \{x \in \mathcal{T}_i : N_S(x) = \mathcal{T}_i\}$. Note that for $x \in I_i$, $\lambda(N_S(x)) \leq \lambda(I_i)$ and $\lambda(N_S(x)) = \lambda(I_i)$ iff $x \in \mathscr{R}_S(N_S)$ where $\lambda(\cdot)$ is the Lebesgue measure on $\mathbb{R}$ (also called as $\mathbb{R}$-Lebesgue measure). So the proximity region of a point in $\mathscr{R}_S(N_S)$ has the largest $\mathbb{R}$-Lebesgue measure. Note that for $\mathcal{Y}_m = \{\mathsf{y}_1, \mathsf{y}_2, \dots, \mathsf{y}_m\} \subset \mathbb{R}$ (i.e. $\Omega = \mathbb{R}$), $\mathscr{R}_S(N_S) = \emptyset$, since $N_S(x) \subseteq I_i$ for all $x \in I_i$ so $N_S(x) \subsetneq \mathbb{R}$ for all $x \in \mathbb{R}$. Note also that given $\mathcal{Y}_m$, $\mathscr{R}_S(N_S)$ is not a random set, but $\mathbf{I}(X \in \mathscr{R}_S(N_S))$ is a random variable.

## 3.1 Vertex and Edge Regions

In $\mathbb{R}$, the spherical proximity maps are defined as open intervals where one of the endpoints is in $\mathcal{Y}_m$. In particular, for $x \in I_i = (\mathsf{y}_{i:m}, \mathsf{y}_{(i+1):m})$ for $i = 1, 2, \dots, m$, $N_S(x) = (\mathsf{y}_{i:m}, \mathsf{y}_{i:m} + 2r(x))$ for all $x \in (\mathsf{y}_{i:m}, \mathsf{y}_{(i-1):m} + \mathsf{y}_{i:m})/2)$ where $r(x) = d(x, \mathsf{y}_{i:m})$ and $N_S(x) = (\mathsf{y}_{(i+1):m}, \mathsf{y}_{(i+1):m} - 2r(x))$ for all $x \in (\mathsf{y}_{(i-1):m} + \mathsf{y}_{i:m})/2, \mathsf{y}_{(i+1):m})$ where $r(x) = d(x, \mathsf{y}_{(i+1):m})$. Hence there are two subinterval in $I_i$ each touching an edge and the midpoint of the interval, and $N_S(x)$ depends on which of these regions $x$ lies in.

In $\mathbb{R}^d$ with $d > 1$, intervals become Delaunay tessellations, and our proximity maps are based on the Delaunay cell $\mathcal{T}_i$ that contains $x$. The region $N_{\mathcal{Y}}(x)$ will also depend on the location of $x$ in $\mathcal{T}_i$ with respect to the vertices or faces (edges in $\mathbb{R}^2$) of $\mathcal{T}_i$. Hence for $N_{\mathcal{Y}}(x)$ to be well-defined, the vertex or face of $\mathcal{T}_i$

associated with $x$ should be uniquely determined. This will give rise to two new concepts: *vertex regions* and face regions (*edge regions* in $\mathbb{R}^2$).

Let $\mathcal{Y}_3 = \{\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3\}$ be three non-collinear points in $\mathbb{R}^2$ and $T(\mathcal{Y}_3) = T(\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3)$ be the triangle with vertices $\mathcal{Y}_3$. To define new proximity regions based on some sort of distance or dissimilarity relative to the vertices $\mathcal{Y}_3$, we associate each point in $T(\mathcal{Y}_3)$ to a vertex of $T(\mathcal{Y}_3)$. This gives rise to the concept of *vertex regions*.

**Definition 3.2.** The connected regions that partition the triangle, $T(\mathcal{Y}_3)$, (in the sense that the pairwise intersections of the regions have zero $\mathbb{R}^2$-Lebesgue measure) such that each region has one and only one vertex of $T(\mathcal{Y}_3)$ on its boundary are called *vertex regions*. $\square$

This definition implies that we have three vertex regions. In fact, we can describe the vertex regions starting with a point $M \in \mathbb{R}^2 \setminus \mathcal{Y}_3$ as follows. Join the point $M$ to a point on each edge by a curve such that the resultant regions satisfy the above definition. We call such regions *$M$-vertex regions* and denote the vertex region associated with vertex $\mathsf{y}$ as $R_M(\mathsf{y})$ for $\mathsf{y} \in \mathcal{Y}_3$. Vertex regions can be defined using any point $M \in \mathbb{R}^2 \setminus \mathcal{Y}_3$ by joining $M$ to a point on each edge. In particular, we use a *center* of $T(\mathcal{Y}_3)$ as the starting point $M$ for vertex regions. See the discussion of triangle centers in (Ceyhan (2009)) with relevant references. We think of the points in $R_M(\mathsf{y})$ as being "closer" to $\mathsf{y}$ than to the other vertices. It is reasonable to require that the area of the region $R_M(\mathsf{y})$ gets larger as $d(M, \mathsf{y})$ increases. Unless stated otherwise, $M$-vertex regions will refer to regions constructed by joining $M$ to the edges with *straight line segments*. Vertex regions with circumcenter, incenter, and center of mass are investigated in Ceyhan (2009). For example $M$-vertex regions can be constructed with $M \in T(\mathcal{Y}_3)^o$ by using the extensions of the line segments joining $\mathsf{y}$ to $M$ for all $\mathsf{y} \in \mathcal{Y}_3$. See Figure 1 (left) with $M = M_C$.

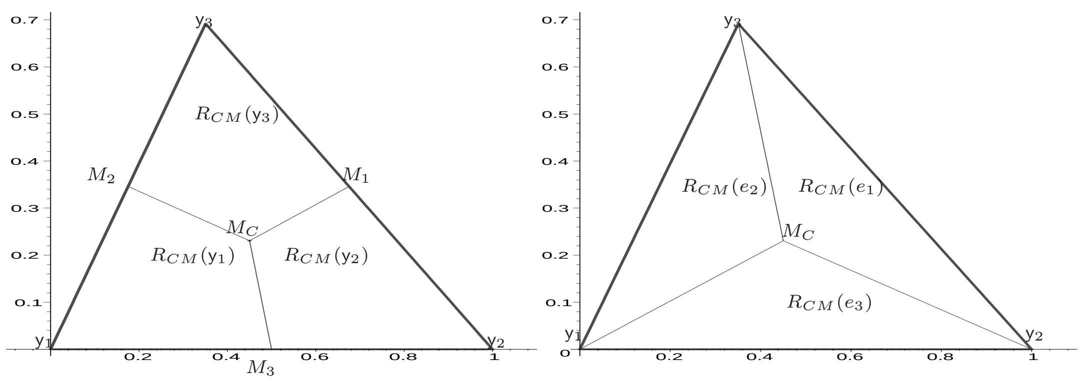


Figure 1: The $CM$-vertex regions (left) and $CM$-edge regions (right) with median lines.

We can also view the endpoints of the interval $I_i$ as the edges of the interval which suggests the concept of *edge regions*.

**Definition 3.3.** The connected regions that partition the triangle, $T(\mathcal{Y}_3)$, in such a way that each region has one and only one edge of $T(\mathcal{Y}_3)$ on its boundary, are called *edge regions*. $\square$

This definition implies that we have exactly three edge regions. In fact, we can describe the edge regions starting with $M$ in $T(\mathcal{Y}_3)^o$, the interior of $T(\mathcal{Y}_3)$. Join the point $M$ to the vertices by curves such that the resultant regions satisfy the above definition. We call such regions *$M$-edge regions* and denote the region for edge $e$ as $R_M(e)$ for $e \in \{e_1, e_2, e_3\}$. Unless stated otherwise, $M$-edge regions will refer to the regions constructed by joining $M$ to the vertices by straight lines. In particular, we use a *center* of $T(\mathcal{Y}_3)$ for the starting point $M$ as the edge regions. See Figure 1 (right) with $M = M_C$. We can also consider the points in $R_M(e)$ to be "closer" to $e$ than to the other edges. Furthermore, it is reasonable to require that the area

of the region $R_M(e)$ gets larger as $d(M,e)$ increases. Moreover, in higher dimensions, the corresponding regions are called "face regions".

Edge regions for incenter, center of mass, and orthocenter are investigated in (Ceyhan (2009)).

# 4 $\Gamma_1$-Regions for Proximity Maps and the Related Concepts

For any set $B \subseteq \Omega$, the $\Gamma_1$-*region* of $B$ associated with $N(\cdot)$, is defined to be the region $\Gamma_1(B,N) := \{z \in \Omega : B \subseteq N(z)\}$. For $x \in \Omega$, we denote $\Gamma_1\big(\{x\},N\big)$ as $\Gamma_1(x,N)$. Note that $\Gamma_1$-region is based on the proximity region $N(\cdot)$. If $\mathcal{X}_n = \big\{X_1, X_2, \dots, X_n\big\}$ is a set of $\Omega$-valued random variables, then $\Gamma_1(X_i,N)$, $i = 1,2,\dots,n$ are random sets. If the $X_i$ are independent and identically distributed, then so are the random sets $\Gamma_1(X_i,N)$. Additionally, $\Gamma_1(\mathcal{X}_n,N)$ is also a random set.

In a digraph $D = (\mathcal{V},\mathcal{A})$, a vertex $v \in \mathcal{V}$ *dominates* itself and all vertices of the form $\{u : (v,u) \in \mathcal{A}\}$. A *dominating set* $S_D$ for the digraph $D$ is a subset of $\mathcal{V}$ such that each vertex $v \in \mathcal{V}$ is dominated by a vertex in $S_D$. A *minimum dominating set* $S^*_D$ is a dominating set of minimum cardinality and the *domination number* $\gamma(D)$ is defined as $\gamma(D) := |S^*_D|$ (see, e.g., Lee (1998)) where $|\cdot|$ denotes the set cardinality functional.

For $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} F$ the domination number of the associated data-random PCD, denoted $\gamma_n(N)$, is the minimum number of points that dominate all points in $\mathcal{X}_n$. Note that, $\gamma_n(N) = 1$ iff $\mathcal{X}_n \cap \Gamma_1(\mathcal{X}_n,N) \neq \emptyset$. Hence the name $\Gamma_1$-*region*. Suppose $\mu$ is a measure on $\Omega$. Following are some general results about $\Gamma_1(\cdot,N)$.

**Proposition 4.1.** *For any proximity map $N$ and set $B \subseteq \Omega$, $\mathscr{R}_S(N) \subseteq \Gamma_1(B,N)$.*

**Proof:** For $x \in \mathscr{R}_S(N)$, $N(x) = \Omega$, so $B \subseteq N(x)$ since $B \subseteq \Omega$. Then $x \in \Gamma_1(B,N)$, hence $\mathscr{R}_S(N) \subseteq \Gamma_1(B,N)$. ■

**Lemma 4.2.** *For any proximity map $N$ and $B \subseteq \Omega$, $\Gamma_1(B,N) = \bigcap_{x\in B} \Gamma_1(x,N)$.*

**Proof:** Given a proximity map $N$ and subset $B \subseteq \Omega$, $y \in \Gamma_1(B,N)$ iff $B \subseteq N(y)$ iff $x \in N(y)$ for all $x \in B$ iff $y \in \Gamma_1(x,N)$ for all $x \in B$ iff $y \in \bigcap_{x\in B} \Gamma_1(x,N)$. Hence the result follows. ■

**Corollary 4.3.** *For any proximity map $N$ and a realization $\mathcal{X}_n = \big\{x_1, x_2, \dots, x_n\big\}$ from $F$ with support $\mathcal{S}(F) \subseteq \Omega$, $\Gamma_1(\mathcal{X}_n,N) = \bigcap_{i=1}^n \Gamma_1(x_i,N)$.*

A problem of interest is finding, if possible, a subset of $B$, say $G \subseteq B$, such that $\Gamma_1(B,N) = \bigcap_{x\in G} \Gamma_1(x,N)$. This implies that only the points in $G$ are used in determining $\Gamma_1(B,N)$.

**Definition 4.4.** An *active set* of points $S_A(B) \subseteq \Omega$ for determining $\Gamma_1(B,N)$ is defined to be a subset of $B$ such that $\Gamma_1(B,N) = \bigcap_{x\in S_A(B)} \Gamma_1(x,N)$. □

This definition allows $B$ to be an active set, which always holds by Lemma 4.2. If $B$ is a set of finitely many points, so is the associated active set. Among the active sets, we seek an active set of minimum cardinality.

**Definition 4.5.** Let $B$ be a set of finitely many points. An active subset of $B \subset \Omega$ is called a *minimal active subset*, denoted $S_\mu(B)$, if there is no other active subset $S_A$ of $B$ such that $S_A(B) \subsetneq S_\mu(B)$. The minimum cardinality among the active subsets of $B$ is called the $\eta$-*value* and denoted as $\eta(B,N)$. An active subset of cardinality $\eta(B,N)$ is called a *minimum active subset* denoted as $S_M(B)$; that is, $\eta(B,N) := |S_M(B)|$. □

Note that Definitions 4.4 and 4.5 can be extended for any subset $B \subseteq \Omega$, in a similar fashion. Moreover, a minimal active set of minimum cardinality is a minimum active set. We will suppress the dependence on $B$ for $S_A(B)$, $S_\mu(B)$, and $S_M(B)$ if there is no ambiguity. In particular, if $B = \mathcal{X}_n$ is a set of $\Omega$-valued random variables, then $S_A$ and $S_M$ are random sets and $\eta_n(N)$ is a random quantity.

For example, in $\mathbb{R}$ with $\mathcal{Y}_2 = \{0, 1\}$, and $\mathcal{X}_n$ a random sample (i.e., set of iid random variables) of size $n > 1$ from $F$ whose support is in $(0, 1)$, $\Gamma_1(\mathcal{X}_n, N_S) = \left(\dfrac{X_{n:n}}{2}, \dfrac{1 + X_{1:n}}{2}\right)$, where $X_{i:n}$ is the $i^{th}$ largest value in $\mathcal{X}_n$. So the extrema (minimum and maximum) of the set $\mathcal{X}_n$ are sufficient to determine the $\Gamma_1$-region; i.e., $S_M = \{X_{1:n}, X_{n:n}\}$. Then $\eta_n(N_S) = 1 + \mathbf{I}(n > 1)$ a.s. for $\mathcal{X}_n$ being a random sample from a continuous distribution with support in $(0, 1)$.

In the multidimensional case there is no natural extension of ordering that yields natural extrema such as minimum or maximum. Some extensions of ordering are proposed under the title of "statistical depth" (see for example Liu et al. (1999)) which is not pursued here. To get the minimum active sets associated with our proximity maps, we will resort to some other type of extrema, such as, the closest points to edges or vertices in $T(\mathcal{Y}_3)$.

For any proximity map $N$ and $\mathcal{X}_n$, $\eta_n(N) \le n$ follows trivially, since

$$\eta_n(N) := \min_{A \subseteq \mathcal{X}_n} \left\{ |A| : \Gamma_1(\mathcal{X}_n, N) = \bigcap_{Z \in A} \Gamma_1(Z) \right\}.$$

**Lemma 4.6.** *Given a sequence of $\Omega$-valued random variables $X_1, X_2, \ldots$ from distribution $F$, let $\mathcal{X}(n) := \mathcal{X}(n-1) \cup \{X_n\}$ for $n = 0, 1, 2, \ldots$ with $\mathcal{X}(0) := \emptyset$. Then $\Gamma_1(\mathcal{X}(n), N)$ is non-increasing in $n$ in the sense that $\Gamma_1(\mathcal{X}(n+1), N) \subseteq \Gamma_1(\mathcal{X}(n), N)$.*

**Proof:** Given a particular type of proximity map $N$ and a data set $\mathcal{X}(n) = \{X_1, X_2, \ldots, X_n\}$, by Lemma 4.2, $\Gamma_1(\mathcal{X}(n), N) = \bigcap_{i=1}^{n} \Gamma_1(X_i, N)$ and by definition, $\mathcal{X}(n+1) = \mathcal{X}(n) \cup \{X_{n+1}\}$. So,

$$\Gamma_1(\mathcal{X}(n+1), N) = \bigcap_{i=1}^{n+1} \Gamma_1(X_i, N) =$$
$$\left[\bigcap_{i=1}^{n} \Gamma_1(X_i, N)\right] \bigcap \Gamma_1(X_{n+1}, N) = \Gamma_1(\mathcal{X}(n), N) \cap \Gamma_1(X_{n+1}, N) \subseteq \Gamma_1(\mathcal{X}(n), N).$$

Thus we have shown that $\Gamma_1(\mathcal{X}(n), N)$ is non-increasing in $n$; i.e., $\Gamma_1(\mathcal{X}(n+1), N) \subseteq \Gamma_1(\mathcal{X}(n), N)$. ■

*Remark* 4.7. By monotone sequential continuity from above (Billingsley (1995)), the sequence $\{\Gamma_1(\mathcal{X}(n), N)\}_{n=1}^{\infty}$ has a limit

$$\begin{aligned} G_1 &:= \bigcap_{j=1}^{\infty} \Gamma_1(\mathcal{X}(j), N) = \lim_{m \to \infty} \bigcap_{j=1}^{m} \Gamma_1(\mathcal{X}(j), N) = \lim_{m \to \infty} \Gamma_1(\mathcal{X}(m), N) \qquad (1) \\ &= \lim_{m \to \infty} \bigcap_{i=1}^{m} \Gamma_1(X_i, N) = \bigcap_{i=1}^{\infty} \Gamma_1(X_i, N). \ \square \end{aligned}$$

**Theorem 4.8.** *Given a sequence of random variables $X_1, X_2, \ldots$ which are identically distributed as $F$ on $\Omega$, let $\mathcal{X}(n) := \mathcal{X}(n-1) \cup \{X_n\}$ with $\mathcal{X}(0) := \emptyset$. Then $\Gamma_1(\mathcal{X}(n), N) \downarrow \mathscr{R}_S(N)$, as $n \to \infty$ a.s. in the sense that $\Gamma_1(\mathcal{X}(n+1), N) \subseteq \Gamma_1(\mathcal{X}(n), N)$ and $\mu(\Gamma_1(\mathcal{X}(n), N) \setminus \mathscr{R}_S(N)) \downarrow 0$ a.s.*

**Proof:** By Lemma 4.6, and by monotone sequential continuity from above, $\{\Gamma_1(\mathcal{X}(n), N)\}_{n=1}^{\infty}$ has a limit, namely, $G_1$ in Equation (1). We claim that $G_1 = \mathscr{R}_S(N)$ a.s.

Suppose $\mathscr{R}_S(N) \subsetneq \Omega$, since if $\mathscr{R}_S(N) = \Omega$ then $N(x) = \Omega$ for all $x \in \Omega$, so $\Gamma_1(\mathcal{X}(n), N) = \Omega$ for all $x$ hence the result would follow trivially. Since $\mathscr{R}_S(N) \subseteq \Gamma_1(\mathcal{X}(n), N)$ for all $n$, $\mathscr{R}_S(N) \subseteq G_1$. From Proposition 5.6. in Karr (1992), $Y_n \xrightarrow{a.s.} Y$ iff $\forall \varepsilon > 0$, $\lim_{n \to \infty} P\left(\sup_{k \ge n} |Y_k - Y| > \varepsilon\right) = 0$. Let $\varepsilon > 0$. Then

$$P\left(\sup_{k \ge n} \mu\left(\bigcap_{i=1}^{k} \Gamma_1(X_i, N) \setminus \mathscr{R}_S(N)\right) \le \varepsilon\right) = P\left(\mu\left(\bigcap_{i=1}^{n} \Gamma_1(X_i, N) \setminus \mathscr{R}_S(N)\right) \le \varepsilon\right) \to 1$$

as $n \to \infty$, because if $\Gamma_1(\mathcal{X}(n), N) \setminus \mathscr{R}_S(N)$ had positive measure, then for each $y \in \Gamma_1(\mathcal{X}(n), N) \setminus \mathscr{R}_S(N)$, $\Omega \setminus N(y)$ will contain data points from $\mathcal{X}(k)$ with positive probability for sufficiently large $k \geq n$. So $y$ can not be in $\Gamma_1(\mathcal{X}(n), N)$, which is a contradiction. Hence the desired result follows. ■

Note however that $\Gamma_1(\mathcal{X}_n, N)$ is neither strictly decreasing nor non-increasing provided that $\mathscr{R}_S(N) \neq \Omega$ for all $\mathcal{X}_n$, because we might have $\Gamma_1(\mathcal{X}_n, N) \subsetneq \Gamma_1(\mathcal{X}_m, N)$ for some $m > n$. Nevertheless, the following two results hold.

**Proposition 4.9.** *Suppose $\Omega \setminus \mathscr{R}_S(N)$ has positive measure. For positive integers $m > n$, let $\mathcal{X}_n\}$ and $\mathcal{X}_m$ be two samples from $F$ on $\Omega$. Then $\mu(\Gamma_1(\mathcal{X}_m, N)) \leq^{ST} \mu(\Gamma_1(\mathcal{X}_n, N))$.*

**Proof:** Recall that for $X \sim F$ and $Y \sim G$, $X \leq^{ST} Y$ if $F(x) \geq G(x)$ for all $x$ with strict inequality holding for at least one $x$.

Let $m > n$ and $\mathcal{X}_n$ and $\mathcal{X}_m$ be two samples from $F$. Then $\mu(\Gamma_1(\mathcal{X}_m, N))$ is more often smaller than $\mu(\Gamma_1(\mathcal{X}_n, N))$. Hence $P[\mu(\Gamma_1(\mathcal{X}_m, N)) \leq \mu(\Gamma_1(\mathcal{X}_n, N))] \geq 1/2$ which only shows stochastic precedence (Boland et al. (2004)).

Now, let $t \in (\mu(\mathscr{R}_S(N)), \mu(\Omega))$, then $\mu(\Gamma_1(\mathcal{X}_m, N)) \leq t$ happens more often than $\mu(\Gamma_1(\mathcal{X}_n, N)) \leq t$, hence $P(\mu(\Gamma_1(\mathcal{X}_m, N)) \leq t) \geq P(\mu(\Gamma_1(\mathcal{X}_n, N)) \leq t)$; that is, $F_m(t) \geq F_n(t)$, where $F_i(\cdot)$ is the distribution function for $\mu(\Gamma_1(\mathcal{X}_i, N))$ for $i = m, n$. For $t < \mu(\mathscr{R}_S(N))$ or $t > \mu(\Omega)$, $F_i(t) = 0$ for $i = m, n$. Letting $N_n := |\mathcal{X}_n \setminus \mathscr{R}_S(N)|$ and $N_m := |\mathcal{X}_m \setminus \mathscr{R}_S(N)|$, then $P(N_n \neq N_m) > 0$ since $\mu(\Omega \setminus \mathscr{R}_S(N)) > 0$. In fact, $P(N_m > N_n) \geq 1/2$. But if $F_m(t) = F_n(t)$ for all $t$ were the case, then $P(N_n = N_m) = 1$ would hold, which is a contradiction. ■

**Theorem 4.10.** *Let $\{\mathcal{X}_n\}_{n=1}^{\infty}$ be a sequence of samples of size $n$ from distribution $F$ with support on $\Omega$. Then $\Gamma_1(\mathcal{X}_n, N) \xrightarrow{p} \mathscr{R}_S(N)$ in the sense that $\mu(\Gamma_1(\mathcal{X}_n, N) \setminus \mathscr{R}_S(N)) \xrightarrow{p} 0$ as $n \to \infty$.*

**Proof:** Given a sequence $\{\mathcal{X}_n\}_{n=1}^{\infty}$ as in the theorem. By Proposition 4.1, $\mathscr{R}_S(N) \subseteq \Gamma_1(\mathcal{X}_n, N)$ for each $n$. If $\Gamma_1(\mathcal{X}_n, N) \setminus \mathscr{R}_S(N)$ has zero measure as $n \to \infty$ then result follows trivially. Otherwise, if $\Gamma_1(\mathcal{X}_n, N) \setminus \mathscr{R}_S(N)$ had positive measure in the limit, for each $y \in \lim_{n\to\infty} \Gamma_1(\mathcal{X}_n, N) \setminus \mathscr{R}_S(N)$, $\Omega \setminus N(y)$ would have positive measure with positive probability, then $\mathcal{X}_n \cap [\Omega \setminus N(y)] \neq \emptyset$ with positive probability for sufficiently large $n$, then $y \notin \Gamma_1(\mathcal{X}_n, N)$, which is a contradiction. ■

Since $\Gamma_1(\mathcal{X}_n, N) = \bigcap_{i=1}^{n} \Gamma_1(X_i, N)$ for a given realization of the data set $\mathcal{X}_n$, first we describe the region $\Gamma_1(X, N)$ for $X \in \mathcal{X}_n$, and then describe the region $\Gamma_1(\mathcal{X}_n, N)$.

**Theorem 4.11.** *If the superset region for any type of proximity map $N$ has positive measure (i.e., $\mu(\mathscr{R}_S(N)) > 0$), then $P(\gamma_n(N) = 1) \to 1$ as $n \to \infty$.*

**Proof:** Notice that if there is at least one data point in $\mathscr{R}_S(N)$ then $\gamma_n(N) = 1$, because any point $x \in \mathscr{R}_S(N)$ will have $N(x) = \Omega$, so $P(\text{there is at least 1 point in } \mathscr{R}_S(N)) \leq P(\gamma_n(N) = 1)$. Now, $P(\text{there is at least 1 point in } \mathscr{R}_S(N)) = 1 - P(\mathcal{X}_n \cap \mathscr{R}_S = \emptyset) = 1 - \left(\dfrac{\mu(\Omega) - \mu(\mathscr{R}_S(N))}{\mu(\Omega)}\right)^n$, which goes to 1 as $n \to \infty$. Hence $P(\gamma_n(N) = 1) \to 1$ as $n \to \infty$. ■

The *relative arc density* of a digraph $D = (\mathcal{V}, \mathcal{A})$ of order $|\mathcal{V}| = n$, denoted as $\rho(D)$, is defined as $\rho(D) = \dfrac{|\mathcal{A}|}{n(n-1)}$ where $|\cdot|$ denotes the cardinality of sets (Janson et al. (2000)). Thus $\rho(D)$ represents the ratio of the number of arcs in the digraph $D$ to the number of arcs in the complete symmetric digraph of order $n$, which is $n(n-1)$.

**Theorem 4.12.** *If support of the joint distribution of $\mathcal{X}_n$ is subset of the superset region for any type of proximity map, then relative arc density $\rho(\mathcal{X}_n) = 1$ a.s.*

**Proof:** Suppose the support of $\mathcal{X}_n$ is a subset of the superset region, then the corresponding digraph is complete with $n(n-1)$ arcs. Hence the relative arc density is 1 with probability 1. ■

# 5 Transformations Preserving Uniformity on Triangles in $\mathbb{R}^2$

The proximity regions and hence the corresponding PCDs are based on the Delaunay tessellation of $\mathcal{Y}_m$, which partitions $C_H(\mathcal{Y}_m)$. So, suppose the set $\mathcal{X}_n$ is a set of iid uniform random variables on the convex hull of $\mathcal{Y}_m$; i.e., a random sample from $\mathcal{U}(\mathcal{C}_H(\mathcal{Y}_m))$. In particular, conditional on $|\mathcal{X}_n \cap T_i| > 0$ being fixed, $\mathcal{X}_n \cap T_i$ will also be a set of iid uniform random variables on $T_i$ for $i \in \{1, 2, \dots, J\}$, where $T_i$ is the $i^{th}$ Delaunay cell and $J$ is the total number of Delaunay cells. Reducing the cell (triangle in $\mathbb{R}^2$) $T_i$ as much as possible while preserving uniformity and the probabilities related to PCDs will simplify the notation and calculations. For simplicity we consider $\mathbb{R}^2$ only.

Let $\mathcal{Y}_3 = \{\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3\} \subset \mathbb{R}^2$ be three non-collinear points and $T(\mathcal{Y}_3)$ be the triangle (including the interior) with vertices $\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3$. Let $X_i \stackrel{iid}{\sim} \mathcal{U}(T(\mathcal{Y}_3))$, the uniform distribution on $T(\mathcal{Y}_3)$, for $i = 1, 2, \dots, n$. The probability density function (pdf) of $\mathcal{U}(T(\mathcal{Y}_3))$ is

$$f(u) = \frac{1}{A(T(\mathcal{Y}_3))}\mathbf{I}(u \in T(\mathcal{Y}_3)),$$

where $A(\cdot)$ is the area functional.

The triangle $T(\mathcal{Y}_3)$ can be carried into the first quadrant by a composition of transformations (scaling, translation, rotation, and reflection) in such a way that the largest edge has unit length and lies on the $x$-axis, and the $x$-coordinate of the vertex nonadjacent to largest edge is less than $1/2$. We call the resultant triangle the *basic triangle* and denote it as $T_b$ where $T_b = \big((0,0),(1,0),(c_1,c_2)\big)$ with $0 < c_1 \le 1/2$, and $c_2 > 0$ and $(1-c_1)^2 + c_2^2 \le 1$. The transformation from any triangle to $T_b$ is denoted by $\phi_b$. See Ceyhan (2009) for a detailed description of $\phi_b$. Notice that $T(\mathcal{Y}_3)$ is transformed into $T_b$, then $T(\mathcal{Y}_3)$ is similar to $T_b$ and $\phi_b(T(\mathcal{Y}_3)) = T_b$. Thus the random variables $X_i \stackrel{iid}{\sim} \mathcal{U}(T(\mathcal{Y}_3))$ transformed along with $T(\mathcal{Y}_3)$ in the described fashion by $\phi_b$ satisfy $\phi_b(X_i) \stackrel{iid}{\sim} \mathcal{U}(T_b)$. So, without loss of generality, we can assume $T(\mathcal{Y}_3)$ to be $T_b$ for uniform data. The functional form of $T_b$ is

$$T_b = \left\{(x,y) \in \mathbb{R}^2 : y \ge 0;\; y \le (c_2\, x)/c_1;\; y \le c_2\,(1-x)/(1-c_1)\right\}.$$

There are other transformations that preserve uniformity of the random variable, but not similarity of the triangles. We only describe the transformation that maps $T_b$ to the standard equilateral triangle, $T_e = T\big((0,0),(1,0),\big(1/2,\sqrt{3}/2\big)\big)$ for exploiting the symmetry in calculations using $T_e$.

Let $\phi_e : (x,y) \to (u,v)$, where $u(x,y) = x + \dfrac{1-2\,c_1}{\sqrt{3}}\, y$ and $v(x,y) = \dfrac{\sqrt{3}}{2\,c_2}\, y$. Then $\mathsf{y}_1$ is mapped to $(0,0)$, $\mathsf{y}_2$ is mapped to $(1,0)$, and $\mathsf{y}_3$ is mapped to $\big(1/2,\sqrt{3}/2\big)$. See also Figure 2.

Note that the inverse transformation is $\phi_e^{-1}(u,v) = \big(x(u,v), y(u,v)\big)$ where $x(u,v) = u - \dfrac{(1-2\,c_1)}{\sqrt{3}}\, v$ and $y(u,v) = \dfrac{2\,c_2}{\sqrt{3}}\, u$. Then the Jacobian is given by

$$J(x,y) \quad = \quad \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 1 & \frac{2\,c_1-1}{\sqrt{3}} \\ 0 & \frac{2\,c_2}{\sqrt{3}} \end{vmatrix} = \frac{2\,c_2}{\sqrt{3}}.$$

So $f_{U,V}(u,v) = f_{X,Y}(\phi_e^{-1}(u,v))\,|J| = \dfrac{4}{\sqrt{3}}\,\mathbf{I}\big((u,v) \in T_e\big)$. Hence uniformity is preserved.

*Remark* 5.1. The probabilities for uniform data in $T(\mathcal{Y}_3)$ involve the ratio of the event region to $A(T(\mathcal{Y}_3))$. If such ratios is not preserved under $\phi_e$, then the probability content for $N$ depends on the geometry of $T(\mathcal{Y}_3)$. In particular, the probability content for uniform data for $N_S$ and $N_{AS}$ depend on the geometry of the triangle, hence is not geometry invariant (Ceyhan (2009)). For example, $P(X \in N_S(Y))$ and $P(X \in N_{AS}(Y,M))$ depends on $(c_1, c_2)$, hence one has to do the computations for all of (uncountably many) of these triangles. Hence we do not investigate $N_S$ and $N_{AS}$ further in this article. $\square$

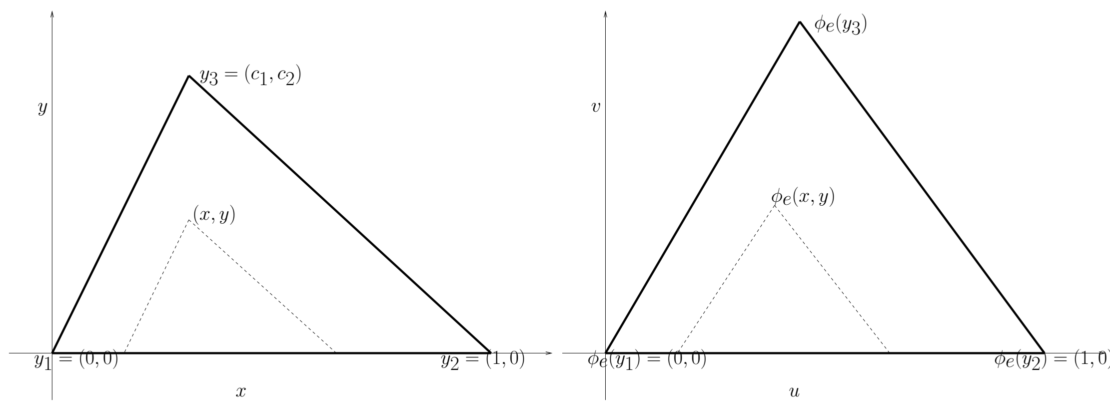


Figure 2: The description of $\phi_e(x, y)$ for $(x, y) \in T_b$ (left) and the equilateral triangle $\phi_e(T_b) = T_e$ (right).

# 6 Sample Proximity Maps in $\mathbb{R}^d$

Let $\mathcal{Y}_m = \{\mathsf{y}_1, \mathsf{y}_2, \ldots, \mathsf{y}_m\}$ be $m$ points in general position in $\mathbb{R}^d$ and $\mathcal{T}_i$ be the $i^{th}$ Delaunay cell for $i = 1, 2, \ldots, J$, where $J$ is the number of Delaunay cells. Let $\mathcal{X}_n$ be a random sample from a distribution $F$ in $\mathbb{R}^d$ with support $\mathcal{S}(F) \subseteq \mathcal{C}_H(\mathcal{Y}_m)$.

In particular, for illustrative purposes, we focus on $\mathbb{R}^2$, where a Delaunay tessellation is a triangulation, provided that no more than three points in $\mathcal{Y}_m$ are cocircular. Furthermore, for simplicity, let $\mathcal{Y}_3 = \{\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3\}$ be three non-collinear points in $\mathbb{R}^2$ and $T(\mathcal{Y}_3) = T(\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3)$ be the triangle with vertices $\mathcal{Y}_3$. Let $\mathcal{X}_n$ be a random sample from $F$ with support $\mathcal{S}(F) \subseteq T(\mathcal{Y}_3)$. In this section, we will describe two families of triangular proximity regions.

## 6.1 Proportional-Edge Proximity Maps

The first type of triangular proximity map we introduce is the proportional-edge proximity map. For this proximity map, the asymptotic distribution of domination number and the relative density of the corresponding PCD will have mathematical tractability.

For $r \in [1, \infty]$, define $N_{PE}^r(\cdot, M) := N(\cdot, M; r, \mathcal{Y}_3)$ to be the *proportional-edge proximity map* with $M$-vertex regions as follows (see also Figure 3 with $M = M_C$ and $r = 2$). For $x \in T(\mathcal{Y}_3) \setminus \mathcal{Y}_3$, let $v(x) \in \mathcal{Y}_3$ be the vertex whose region contains $x$; i.e., $x \in R_M(v(x))$. If $x$ falls on the boundary of two $M$-vertex regions, we assign $v(x)$ arbitrarily. Let $e(x)$ be the edge of $T(\mathcal{Y}_3)$ opposite $v(x)$. Let $\ell(v(x), x)$ be the line parallel to $e(x)$ through $x$. Let $d(v(x), \ell(v(x), x))$ be the Euclidean (perpendicular) distance from $v(x)$ to $\ell(v(x), x)$. For $r \in [1, \infty)$, let $\ell_r(v(x), x)$ be the line parallel to $e(x)$ such that

$$d(v(x), \ell_r(v(x), x)) = r\, d(v(x), \ell(v(x), x)) \text{ and } d(\ell(v(x), x), \ell_r(v(x), x)) < d(v(x), \ell_r(v(x), x)).$$

Let $T_r(x)$ be the triangle similar to and with the same orientation as $T(\mathcal{Y}_3)$ having $v(x)$ as a vertex and $\ell_r(v(x), x)$ as the opposite edge. Then the proportional-edge proximity region $N_{PE}^r(x, M)$ is defined to be $T_r(x) \cap T(\mathcal{Y}_3)$. Notice that $\ell(v(x), x)$ divides the edges of $T_r(x)$ (other than $\ell_r(v(x), x)$) proportionally with the factor $r$. Hence the name *proportional edge proximity region*.

Notice that $r \geq 1$ implies $x \in N_{PE}^r(x, M)$. Furthermore, $\lim_{r\to\infty} N_{PE}^r(x, M) = T(\mathcal{Y}_3)$ for all $x \in T(\mathcal{Y}_3) \setminus \mathcal{Y}_3$, so we define $N_{PE}^\infty(x, M) = T(\mathcal{Y}_3)$ for all such $x$. For $x \in \mathcal{Y}_3$, we define $N_{PE}^r(x, M) = \{x\}$ for all $r \in [1, \infty]$.

Notice that $X_i \stackrel{iid}{\sim} F$, with the additional assumption that the non-degenerate two-dimensional probability density function $f$ exists with support $\mathcal{S}(F) \subseteq T(\mathcal{Y}_3)$, implies that the special case in the construction of

$N_{PE}^r$ — $X$ falls on the boundary of two vertex regions — occurs with probability zero. Note that for such an $F$, $N_{PE}^r(X)$ is a triangle a.s.

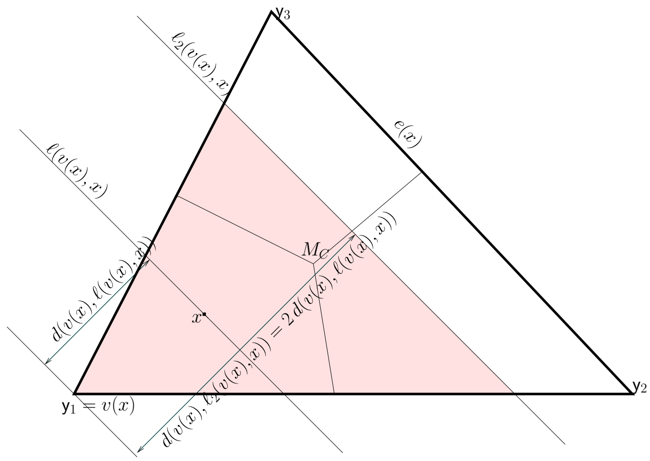


Figure 3: Construction of proportional edge proximity region, $N_{PE}^2(x)$ (shaded region).

The functional form of $N_{PE}^r(x, M)$ for $x = (x_0, y_0) \in T_b$ is given in Ceyhan (2009). Of particular interest is $N_{PE}^r(x, M)$ with any $M$ and $r \in \{\sqrt{2},\, 3/2, 2\}$. For $r = \sqrt{2}$, $\ell(v(x), x)$ divides $T_{\sqrt{2}}(x)$ into two regions of equal area, hence $N_{PE}^{\sqrt{2}}(x, M)$ is also referred to as *double-area proximity region*. For $r = 2$, $\ell(v(x), x)$ divides the edges of $T_2(x)$ —other than $\ell_r(v(x), x)$ — into two segments of equal length, hence $N_{PE}^2(x, M)$ is also referred to as *double-edge proximity region*. For $r < 3/2$, $\mathscr{R}_S(N_{PE}^r, M_C) = \emptyset$, and for $r > 3/2$, $\mathscr{R}_S(N_{PE}^r, M_C)$ has positive area; for $r = 3/2$, $\mathscr{R}_S(N_{PE}^r, M_C) = \{M_C\}$. Therefore, $r = 3/2$ is the threshold for the superset region to be nonempty. Furthermore, $r = 3/2$ will be the value at which the asymptotic distribution of the domination number of the PCD based on $N_{PE}^r(\cdot, M_C)$ is nondegenerate (Ceyhan and Priebe (2005)).

Let $\mathscr{R}_S^{\perp}\left(N_{PE}^r, M\right)$ be the superset region for $N_{PE}^r$ based on $M$-vertex regions with orthogonal projections. Then the superset region with the incenter $\mathscr{R}_S^{\perp}\left(N_{PE}^2, M_I\right)$ is as in Figure 4 (right). Let $M_i$ be the midpoint of edge $e_i$ for $i = 1, 2, 3$. Then $T(M_1, M_2, M_3) \subseteq \mathscr{R}_S^{\perp}(N_{PE}^2, M_I)$ for all $T(\mathcal{Y}_3)$ with equality holding when $T(\mathcal{Y}_3)$ is an equilateral triangle.

For $N_{PE}^2(\cdot, M_C)$ constructed using the median lines $\mathscr{R}_S\left(N_{PE}^2, M_C\right) = T(M_1, M_2, M_3)$, and for $N_{PE}^2(\cdot, M_C)$ constructed by the orthogonal projections, $\mathscr{R}_S^{\perp}\left(N_{PE}^2, M_C\right) \supseteq T(M_1, M_2, M_3)$ with equality holding when

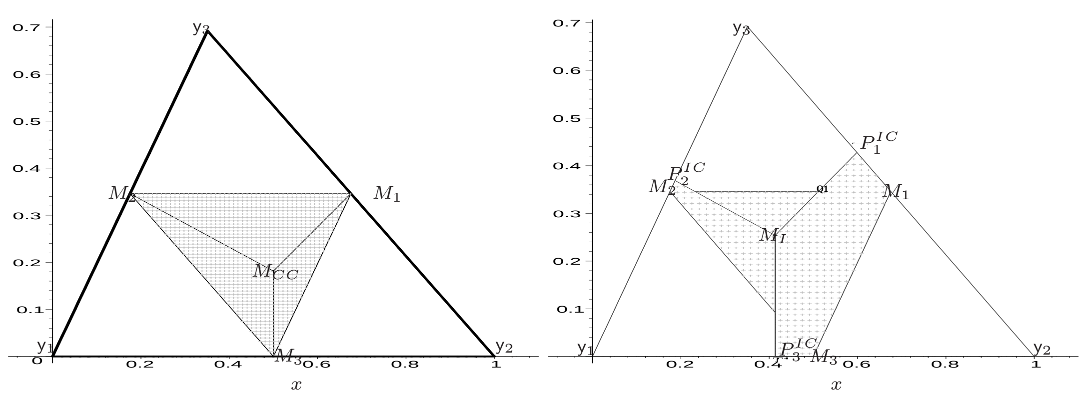


Figure 4: Superset region $\mathscr{R}_S\left(N_{PE}^2, M_{CC}\right)$ in an acute triangle (left), superset region, $\mathscr{R}_S^{\perp}(N_{PE}^2, M_I)$ (right)

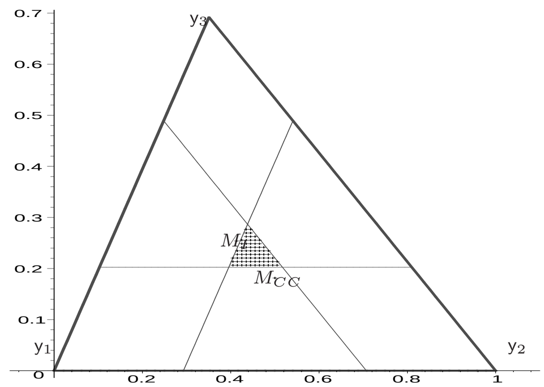


Figure 5: The triangle $\mathscr{T}^{r=\sqrt{2}}$.

$T(\mathcal{Y}_3)$ is an equilateral triangle.

In $T(\mathcal{Y}_3)$, drawing the lines $\xi_i(r,x)$ such that $d(\mathsf{y}_i, e_i) = r\, d(\xi_i(r,x), \mathsf{y}_i)$ for $j \in \{1,2,3\}$ yields a triangle, $\mathscr{T}^r$, for $r < 3/2$ . See Figure 5 for $\mathscr{T}^r$ with $r = \sqrt{2}$.

The functional form of $\mathscr{T}^r$ in $T_b$ is

$$\begin{aligned}\mathscr{T}^r &= \left\{(x,y) \in T_b : y \geq \frac{c_2\,(r-1)}{r};\ y \leq \frac{c_2\,(1-r\,x)}{r\,(1-c_1)};\ y \leq \frac{c_2\,(r\,(x-1)+1)}{r\,c_1}\right\} \\ &= T\left(\left(\frac{(r-1)\,(1+c_1)}{r}, \frac{c_2\,(r-1)}{r}\right), \left(\frac{2-r+c_1\,(r-1)}{r}, \frac{c_2\,(r-1)}{r}\right), \left(\frac{c_1\,(2-r)+r-1}{r}, \frac{c_2\,(r-2)}{r}\right)\right).\end{aligned} \tag{2}$$

There is a crucial difference between $\mathscr{T}^r$ and $T(M_1,M_2,M_3)$: $T(M_1,M_2,M_3) \subseteq \mathscr{R}_S\,(N^r_{PE}, M)$ for all $M$ and $r \geq 2$, but $(\mathscr{T}^r)^o$ and $\mathscr{R}_S\,(N^r_{PE}, M)$ are disjoint for all $M$ and $r$.

So if $M \in (\mathscr{T}^r)^o$, then $\mathscr{R}_S\,(N^r_{PE}, M) = \emptyset$; if $M \in \partial(\mathscr{T}^r)$, then $\mathscr{R}_S\,(N^r_{PE}, M) = \{M\}$; and if $M \notin \mathscr{T}^r$, then $\mathscr{R}_S\,(N^r_{PE}, M)$ has positive area. The triangle $\mathscr{T}^r$ defined above plays a crucial role in the analysis of the distribution of the domination number of the PCD based on proportional-edge proximity maps. The superset region $\mathscr{R}_S\,(N^r_{PE}, M)$ will be important for both the domination number and the relative density of the corresponding PCDs.

The functional forms of the superset region, $\mathscr{R}_S\,(N^r_{PE}, M)$, and $T(M_1,M_2,M_3)$ in $T_b$ are provided in Ceyhan (2009).

$N^r_{PE}(\cdot, M)$ is geometry invariant if $M$-vertex regions are constructed with $M \in T(\mathcal{Y}_3)^o$ by using the extensions of the line segments joining $\mathsf{y}$ to $M$ for all $\mathsf{y} \in \mathcal{Y}_3$. But when the vertex regions are constructed by orthogonal projections, $N^r_{PE}(\cdot, M)$ is not geometry invariant (Ceyhan (2009)), hence such vertex regions are not considered here.

### 6.1.1 Extension of $N^r_{PE}$ to Higher Dimensions

The extension to $\mathbb{R}^d$ for $d > 2$ is straightforward. The extension with $M = M_C$ is given her, but the extension for general $M$ is similar. Let $\mathcal{Y}_{d+1} = \{\mathsf{y}_1, \mathsf{y}_2, \ldots, \mathsf{y}_{d+1}\}$ be $d+1$ points that do not lie on the same $d-1$-dimensional hyperplane. Denote the simplex formed by these $d+1$ points as $\mathfrak{S}(\mathcal{Y}_{d+1})$. A simplex is the simplest polytope in $\mathbb{R}^d$ having $d+1$ vertices, $d\,(d+1)/2$ edges and $d+1$ faces of dimension $(d-1)$. For $r \in [1, \infty]$, define the proximity map as follows. Given a point $x$ in $\mathfrak{S}(\mathcal{Y}_{d+1})$, let $v := \operatorname{argmin}_{\mathsf{y} \in \mathcal{Y}_{d+1}} \mathrm{V}(Q_{\mathsf{y}}(x))$ where $Q_{\mathsf{y}}(x)$ is the polytope with vertices being the $d\,(d+1)/2$ midpoints of the edges, the vertex $v$ and $x$ and $\mathrm{V}(\cdot)$ is the $d$-dimensional volume functional. That is, the vertex region for vertex $v$ is the polytope with vertices given by $v$ and the midpoints of the edges. Let $v(x)$ be the vertex in whose region $x$ falls. If

$x$ falls on the boundary of two vertex regions, $v(x)$ is assigned arbitrarily. Let $\varphi(x)$ be the face opposite to vertex $v(x)$, and $\Upsilon(v(x), x)$ be the hyperplane parallel to $\varphi(x)$ which contains $x$. Let $d(v(x), \Upsilon(v(x), x))$ be the (perpendicular) Euclidean distance from $v(x)$ to $\Upsilon(v(x), x)$. For $r \in [1, \infty)$, let $\Upsilon_r(v(x), x)$ be the hyperplane parallel to $\varphi(x)$ such that

$$d(v(x), \Upsilon_r(v(x), x)) = r\, d(v(x), \Upsilon(v(x), x)) \text{ and } d(\Upsilon(v(x), x), \Upsilon_r(v(x), x)) < d(v(x), \Upsilon_r(v(x), x)).$$

Let $\mathfrak{S}_r(x)$ be the polytope similar to and with the same orientation as $\mathfrak{S}$ having $v(x)$ as a vertex and $\Upsilon_r(v(x), x)$ as the opposite face. Then the proximity region $N^r_{PE}(x, M_C) := \mathfrak{S}_r(x) \cap \mathfrak{S}(\mathcal{Y}_{d+1})$. Notice that $r \geq 1$ implies $x \in N^r_{PE}(x, M_C)$.

### 6.1.2 $\Gamma_1$-Regions for Proportional-Edge Proximity Maps

For $N^r_{PE}(\cdot, M)$, the $\Gamma_1$-region is constructed as follows; see also Figure 6. Let $\xi_i(r, x)$ be the line parallel to $e_i$ such that $\xi_i(r, x) \cap T(\mathcal{Y}_3) \neq \emptyset$ and $r\, d(\mathsf{y}_i, \xi_i(r, x)) = d(\mathsf{y}_i, \ell(\mathsf{y}_i, x))$ for $i \in \{1, 2, 3\}$. Then

$$\Gamma_1\left(x, N^r_{PE}, M\right) = \bigcup_{i=1}^{3} \left[\Gamma_1\left(x, N^r_{PE}, M\right) \cap R_M(\mathsf{y}_i)\right]$$

where

$$\Gamma_1\left(x, N^r_{PE}, M\right) \cap R_M(\mathsf{y}_i) = \{z \in R_M(\mathsf{y}_i) : d(\mathsf{y}_i, \ell(\mathsf{y}_i, z)) \geq d(\mathsf{y}_i, \xi_i(r, x)\} \text{ for } i \in \{1, 2, 3\}.$$

Notice that $r \geq 1$ implies $x \in \Gamma_1\left(x, N^r_{PE}, M\right)$. Furthermore, $\lim_{r \to \infty} \Gamma_1\left(x, N^r_{PE}, M\right) = T(\mathcal{Y}_3)$ for all $x \in T(\mathcal{Y}_3) \setminus \mathcal{Y}_3$ and so we define $\Gamma_1\left(x, N^{r=\infty}_{PE}, M\right) = T(\mathcal{Y}_3)$ for all such $x$. For $x \in \mathcal{Y}_3$, $\Gamma_1\left(x, N^r_{PE}, M\right) = \{x\}$ for all $r \in [1, \infty]$.

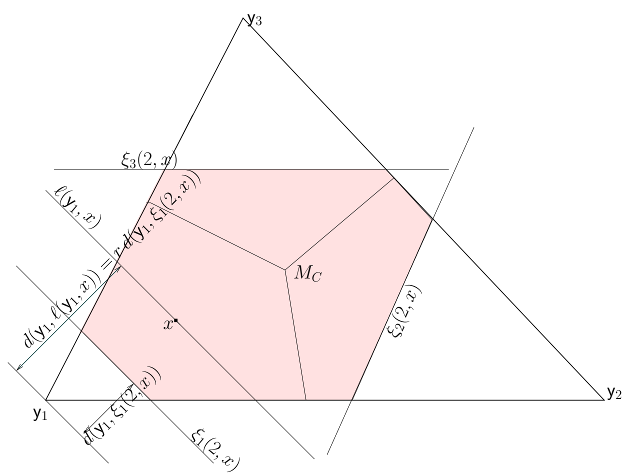


Figure 6: Construction of the $\Gamma_1$-region, $\Gamma_1\left(x, N^{r=2}_{PE}, M_C\right)$ (shaded region).

The functional form of $\Gamma_1\left(x = (x_0, y_0), N^r_{PE}, M\right)$ in the basic triangle $T_b$ is given by

$$\Gamma_1\left(x = (x_0, y_0), N^r_{PE}, M\right) = \bigcup_{i=1}^{3} \left[\Gamma_1\left(x = (x_0, y_0), N^r_{PE}, M\right) \cap R_M(y_i)\right]$$

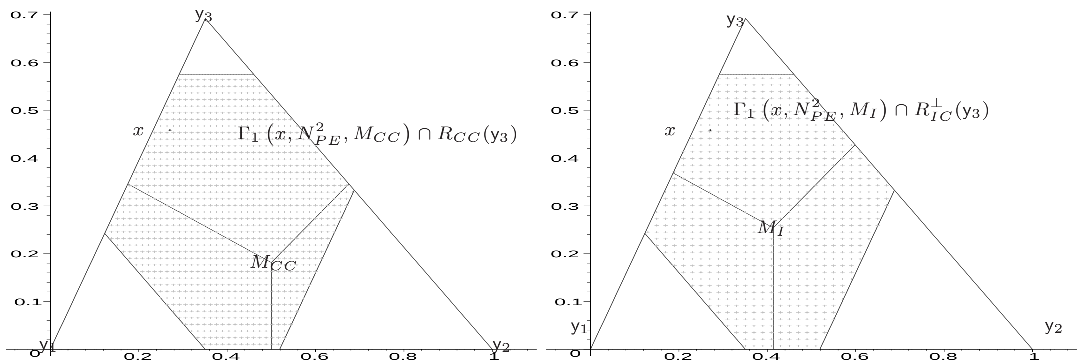


Figure 7: $\Gamma_1\left(x, N_{PE}^2, M_{CC}\right)$ with $x \in R_{CC}(\mathsf{y}_3)$ (left), $x \in R_{IC}^{\perp}(\mathsf{y}_3)$ (right).

where

$$\Gamma_1\left(x = (x_0, y_0), N_{PE}^r, M\right) \cap R_M(\mathsf{y}_1) = \left\{(x, y) \in R_M(\mathsf{y}_1) : y \geq \frac{y_0}{r} - \frac{c_2\,(r\,x - x_0)}{(1 - c_1)\,r}\right\},$$

$$\Gamma_1\left(x = (x_0, y_0), N_{PE}^r, M\right) \cap R_M(\mathsf{y}_2) = \left\{(x, y) \in R_M(\mathsf{y}_1) : y \geq \frac{y_0}{r} - \frac{c_2\,(r\,(x - 1) + 1 - x_0)}{c_1\,r}\right\},$$

$$\Gamma_1\left(x = (x_0, y_0), N_{PE}^r, M\right) \cap R_M(\mathsf{y}_3) = \left\{(x, y) \in R_M(\mathsf{y}_1) : y \leq \frac{y_0 - c_2\,(1 - r)}{r}\right\}.$$

Notice that $\Gamma_1\left(x, N_{PE}^r, M_C\right)$ is a convex hexagon for all $r \geq 2$ and $x \in T(\mathcal{Y}_3) \setminus \mathcal{Y}_3$, (since for such an $x$, $\Gamma_1\left(x, N_{PE}^r, M_C\right)$ is bounded by $\xi_i(r, x)$ and $e_i$ for all $i \in \{1, 2, 3\}$, see also Figure 6) else it is either a convex hexagon or a non-convex polygon depending on the location of $x$ and the value of $r$.

Furthermore, in $\mathbb{R}^d$ with $d > 2$ let $\zeta_i(x)$ be the hyperplane such that $\zeta_i(x) \cap \mathfrak{S}\left(\mathcal{Y}_m\right) \neq \emptyset$ and $r\, d(\mathsf{y}_i, \zeta_i(x)) = d(\mathsf{y}_i, \eta(\mathsf{y}_i, x))$ for $i = 1, 2, \ldots, (d+1)$. Then $\Gamma_1\left(x, N_{PE}^r\right) \cap R_{CM}(\mathsf{y}_i) = \{z \in R_{CM}(\mathsf{y}_i) : d(\mathsf{y}_i, \eta(\mathsf{y}_i, z)) \geq d(\mathsf{y}_i, \zeta_i(x)\}$, for $i = 1, 2, 3$. Hence $\Gamma_1\left(x, N_{PE}^r\right) = \bigcup_{i=1}^{d+1}(\Gamma_1\left(x, N_{PE}^r\right) \cap R_{CM}(\mathsf{y}_i))$. Notice that $r \geq 1$ implies $x \in N_{PE}^r(x)$ and $x \in \Gamma_1\left(x, N_{PE}^r\right)$.

So far, we have described the $\Gamma_1$-region for a point in $x \in T(\mathcal{Y}_3)$. For a set $\mathcal{X}_n$ of size $n$ in $T(\mathcal{Y}_3)$, the region $\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right)$ can be exactly described by the edge extrema.

**Definition 6.1.** The (closest) *edge extrema* of a set $B$ in $T(\mathcal{Y}_3)$ are the points closest to the edges of $T(\mathcal{Y}_3)$, denoted $x_{e_i}$ for $i \in \{1, 2, 3\}$; that is, $x_{e_i} \in \operatorname{arginf}_{x \in B} d(x, e_i)$.

Note that if $B = \mathcal{X}_n$ is a random sample of size $n$ from $F$ then the edge extrema, denoted $X_{e_i}(n)$, are random variables.

**Proposition 6.2.** *Let $B$ be any set of $n$ distinct points in $T(\mathcal{Y}_3)$ and $x_{e_i} \in \operatorname{arginf}_{x \in B} d(x, e_i)$. For proportional-edge proximity maps with $M$-vertex regions, $\Gamma_1\left(B, N_{PE}^r, M\right) = \bigcap_{i=1}^{3} \Gamma_1\left(x_{e_i}, N_{PE}^r, M\right)$.*

**Proof:** Given $B = \{x_1, x_2, \ldots, x_n\}$ in $T(\mathcal{Y}_3)$. Note that

$$\Gamma_1\left(B, N_{PE}^r, M\right) \cap R_M(\mathsf{y}_i) = \left[\bigcap_{i=1}^{n} \Gamma_1\left(x_i, N_{PE}^r, M\right)\right] \cap R_M(\mathsf{y}_i),$$

and if $d(\mathsf{y}_i, \ell(\mathsf{y}_i, x)) \leq d(\mathsf{y}_i, \ell(\mathsf{y}_i, x'))$ then $N_{PE}^r(x, M) \subseteq N_{PE}^r(x', M)$ for all $x, x' \in R_M(y_i)$. Further, by definition $x_{e_i} \in \operatorname{argmax}_{x \in B} d(\mathsf{y}_i, \xi_i(r, x))$, so

$$\Gamma_1\left(B, N_{PE}^r, M\right) \cap R_M(\mathsf{y}_i) = \Gamma_1\left(x_{e_i}, N_{PE}^r, M\right) \cap R_M(\mathsf{y}_i) \text{ for } i \in \{1, 2, 3\}.$$

Furthermore, $\Gamma_1\left(B, N_{PE}^r, M\right) = \bigcup_{i=1}^{3}\left[\Gamma_1\left(x_{e_i}, N_{PE}^r, M\right) \cap R_M(\mathsf{y}_i)\right]$, and

$$\Gamma_1\left(x_{e_i}, N_{PE}^r, M\right) \cap R_M(\mathsf{y}_i) = \left[\bigcap_{j=1}^{3} \Gamma_1\left(x_{e_j}, N_{PE}^r, M\right)\right] \cap R_M(\mathsf{y}_i) \text{ for } i \in \{1,2,3\}.$$

Combining these two results, we obtain $\Gamma_1\left(B, N_{PE}^r, M\right) = \bigcap_{j=1}^{3} \Gamma_1\left(x_{e_j}, N_{PE}^r, M\right)$. ■

From the above proposition, we see that the $\Gamma_1$-region for $B$ as in the proposition can also be written as the union of three regions of the form

$$\Gamma_1\left(B, N_{PE}^r, M\right) \cap R_M(\mathsf{y}_i) = \{z \in R_M(\mathsf{y}_i) : \; d(\mathsf{y}_i, \ell(\mathsf{y}_i, z)) \geq d(\mathsf{y}_i, \xi_i(r, x_{e_i}))\} \text{ for } i \in \{1,2,3\}.$$

**Corollary 6.3.** *Let $\mathcal{X}_n$ be a random sample from a continuous distribution $F$ on $T(\mathcal{Y}_3)$. For proportional-edge proximity maps with $M$-vertex regions, $\eta_n\left(N_{PE}^r\right) \leq 3$ with equality holding with positive probability for $n \geq 3$.*

**Proof:** From Proposition 6.2, $\eta_n\left(N_{PE}^r\right) \leq 3$. Furthermore, $X_e(n)$ is unique for each edge $e$ a.s. since $F$ is continuous, and there are three distinct edge extrema with positive probability. Hence $P(\eta_n\left(N_{PE}^r\right) = 3) > 0$ for $n \geq 3$. ■

Then $\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right) = \bigcap_{i=1}^{3} \Gamma_1\left(X_{e_i}, N_{PE}^r\right)$, where $e_i$ is the edge opposite vertex $\mathsf{y}_i$, for $i = 1,2,3$. So $\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right) \cap R_M(\mathsf{y}_i) = \{z \in R_M(\mathsf{y}_i) : \; d(\mathsf{y}_i, \ell(\mathsf{y}_i, z)) \geq d(\mathsf{y}_i, \xi_i(r, x_{e_i}))\}$, for $i = 1,2,3$.

Note that $P(\eta_n\left(N_{PE}^r\right) = 3) \to 1$ as $n \to \infty$ for $\mathcal{X}_n$ a random sample from $\mathcal{U}(T(\mathcal{Y}_3))$, since edge extrema are distinct with probability 1 as $n \to \infty$ as shown in the following theorem.

**Theorem 6.4.** *Let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(T(\mathcal{Y}_3))$ and let $E_{c,3}(n)$ be the event that (closest) edge extrema are distinct. Then $P(E_{c,3}(n)) \to 1$ as $n \to \infty$.*

**Proof:** Using the uniformity preserving transformation $\phi_e$ without loss of generality, one can assume $\mathcal{X}_n$ is a random sample from $\mathcal{U}(T_e)$. Observe also that the edge extrema in $T_b$ are mapped into the edge extrema in $T_e$. Note that the probability of edge extrema all being equal to each other is $P(X_{e_1}(n) = X_{e_2}(n) = X_{e_3}(n)) = \mathbf{I}(n = 1)$. Let $E_{c,2}(n)$ be the event that there are only two distinct (closest) edge extrema. Then for $n > 1$,

$$P(E_{c,2}(n)) = P(X_{e_1}(n) = X_{e_2}(n)) + P(X_{e_1}(n) = X_{e_3}(n)) + P(X_{e_2}(n) = X_{e_3}(n))$$

since the intersection of events $X_{e_1}(n) = X_{e_2}(n))$, $X_{e_1}(n) = X_{e_3}(n)$, and $X_{e_2}(n) = X_{e_3}(n)$ is equivalent to the event $X_{e_1}(n) = X_{e_2}(n) = X_{e_3}(n)$. Notice also that $P(E_{c,2}(n = 2)) = 1$. So, for $n > 2$, there are two or three distinct edge extrema with probability 1; i.e., $P(E_{c,3}(n)) + P(E_{c,2}(n)) = 1$ for $n > 2$.

We will show that $P(E_{c,2}(n)) \to 0$ as $n \to \infty$, which will imply the desired result.

First consider $P(X_{e_1}(n) = X_{e_2}(n))$. The event $X_{e_1}(n) = X_{e_2}(n) = X_e = (X, Y)$ is equivalent to the event that $\mathcal{X}_n \subset \{U \in T_e : \; d(\mathsf{y}_1, \ell(\mathsf{y}_1, U)) \leq d(\mathsf{y}_1, \ell(\mathsf{y}_1, X_e)), \; d(\mathsf{y}_2, \ell(\mathsf{y}_2, U)) \leq d(\mathsf{y}_2, \ell(\mathsf{y}_2, X_e))\}$. For example, if given $X_{e_1}(n) = X_{e_2}(n) = (x, y)$ the remaining $n - 1$ points will lie in the shaded region in Figure 8 (left). For other pairs of edge extrema, see Figures 8 (right) and 9.

The pdf of such $X_e = (X, Y)$ is $f(x, y) = n\,\left(4/\sqrt{3}\right)\,\left(y^2/\sqrt{3}\right)^{n-1}$. Let $\varepsilon > 0$, by Markov's inequality, $P\left(\sqrt{3}/2 - Y > \varepsilon\right) \leq \mathbf{E}\left[\sqrt{3}/2 - Y\right]/\varepsilon$. But,

$$\begin{aligned}
\mathbf{E}\left[\sqrt{3}/2 - Y\right] &= \int_0^{1/2}\int_0^{\sqrt{3}\,x} \left(\sqrt{3}/2 - y\right)\, n\, \left(y^2/\sqrt{3}\right)^{n-1} \left(4/\sqrt{3}\right) dy\, dx \\
&\quad + \int_{1/2}^{1}\int_0^{\sqrt{3}\,(1-x)} \left(\sqrt{3}/2 - y\right)\, n\, \left(y^2/\sqrt{3}\right)^{n-1} \left(4/\sqrt{3}\right) dy\, dx \\
&= 4\left(\sqrt{3}/4\right)^n \frac{1}{n\,(4\,n^2 - 1)},
\end{aligned}$$

which converges to 0 as $n \to \infty$. So if $X_{e_1}(n) = X_{e_2}(n)$ were the case, then geometric locus of this point goes to $\mathsf{y}_3$. That is, for each $\varepsilon > 0$, $P\left(\sqrt{3}/2 - Y > \varepsilon\right) \to 0$ as $n \to \infty$. Hence $P(X_{e_1}(n) = X_{e_2}(n) \neq \mathsf{y}_3) \to 0$ as $n \to \infty$. Furthermore, $P(X_{e_1}(n) = X_{e_2}(n) = \mathsf{y}_3) \leq P(X_{e_2}(n) \in e_2) = 0$ for all $n \geq 1$. So $P(X_{e_1}(n) = X_{e_2}(n)) \to 0$ as $n \to \infty$.

Likewise, by symmetry, it follows that $\lim_{n\to\infty} P(X_{e_1}(n) = X_{e_3}(n)) = \lim_{n\to\infty} P(X_{e_2}(n) = X_{e_3}(n)) = 0$. Hence $P(E_{c,2}(n)) \to 0$ as $n \to \infty$. Thus $P(E_{c,3}(n)) \to 1$ as $n \to \infty$. ■

The above theorem implies that the asymptotic distribution of $\eta_n\left(N^r_{PE}\right)$ is degenerate with $P(\eta_n\left(N^r_{PE}\right) = 3) \to 1$ as $n \to \infty$. But for finite $n$, $\eta_n\left(N^r_{PE}\right)$ for $X_i \overset{iid}{\sim} \mathcal{U}(T(\mathcal{Y}_3))$ has the following non-degenerate distribution.

$$\eta_n\left(N^r_{PE}\right) = \begin{cases} 2 & \text{wp } \pi_2(n) \\ 3 & \text{wp } \pi_3(n) = 1 - \pi_2(n), \end{cases}$$

where $\pi_2(n) \in (0,1)$ is the probability of edge extrema for any two distinct edges being concurrent.

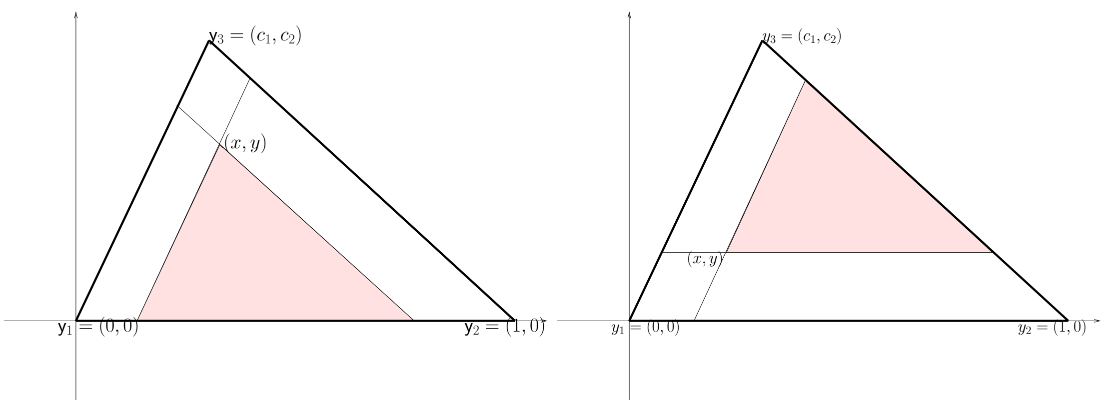


Figure 8: The figure for $X_{e_1}(n) = X_{e_2}(n) = (x,y)$ (left) and $X_{e_2}(n) = X_{e_3}(n) = (x,y)$ (right).

*Remark* 6.5. If $\mathcal{X}_n$ is a random sample from $F$ such that $\mathcal{S}(F) \cap \{x \in T(\mathcal{Y}_3) : d(x, e_i) \leq \varepsilon_1\}$ has positive measure and $\mathcal{S}(F) \cap B(\mathsf{y}_i, \varepsilon_2) = \emptyset$ for some $\varepsilon_1, \varepsilon_2 > 0$, then $P(E_{c,3}(n)) \to 1$ as $n \to \infty$ follows trivially. However, the case that $F$ has positive density around the vertices $\mathcal{Y}_3$ requires more work to prove as shown below. □

**Theorem 6.6.** *Let $\mathcal{X}_n$ be a random sample from $F$ such that $B(\mathsf{y}_i, \varepsilon) \subseteq \mathcal{S}(F)$ for some $\varepsilon > 0$ and for each $i = 1, 2, 3$, then $P(E_{c,3}(n)) \to 1$ as $n \to \infty$.*

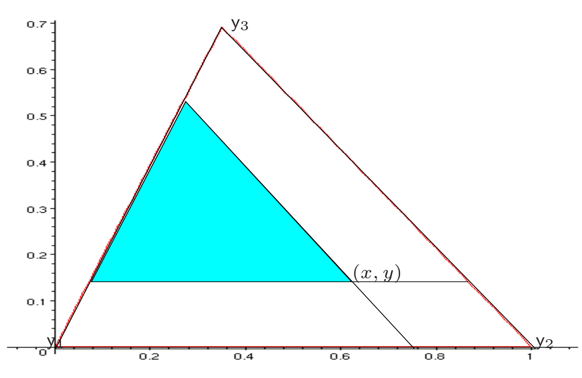


Figure 9: The figure for $X_{e_1}(n) = X_{e_3}(n) = (x,y)$.

**Proof:** Using the transformation $\phi_e : (x,y) \rightarrow (u,v)$, above, we can without loss of generality assume $\mathcal{X}_n$ is a random sample from $F$ with support $\mathcal{S}(F) \subseteq T_e$. After $\phi_e$ is applied, suppose $F$ becomes $F_e$, then $\phi_e(\mathcal{X}_n)$ becomes a random sample from $F_e$ such that $B(\phi_e(\mathsf{y}_i), \varepsilon_e) \subseteq \mathcal{S}(F_e)$ for some $\varepsilon_e > 0$ and for each $i = 1,2,3$. First consider $P(X_{e_1} = X_{e_3})$. Given $X_{e_1} = X_{e_3} = (x,y)$ the remaining $n-1$ points will lie in the shaded region in Figure 8 (right). $X_{e_1} = X_{e_3} = (X,Y)$ is equivalent to the event that

$$\mathcal{X}_n \subset S_R(X,Y) := \{(U,W) \in T_e : \ \ell(\mathsf{y}_1,(U,W)) \leq \ell(\mathsf{y}_1,(X,Y)), \ \ell(\mathsf{y}_3,(U,W)) \leq \ell(\mathsf{y}_3,(X,Y))\}.$$

The pdf of such $(X,Y)$ is $f(x,y) = n\, G(x,y)^{n-1} f(x,y)$ where $G(u,v) := P_F(X \in S_R(u,v))$. Note that $P(X_{e_2} = X_{e_3} = \mathsf{y}_1) = 0$ for all $n$, and $X_{e_2} = X_{e_3} \neq \mathsf{y}_1$ is equivalent to $d((X,Y),\mathsf{y}_1) > 0$. Let $\varepsilon > 0$, by Markov's inequality, $P(d((X,Y),\mathsf{y}_1) > \varepsilon) \leq \mathbf{E}\left[d((X,Y),\mathsf{y}_1)\right]/\varepsilon = \mathbf{E}\left[\sqrt{X^2+Y^2}\right]/\varepsilon$. Switching to the polar coordinates as $X = R\cos\theta$ and $Y = R\sin\theta$, we get $\sqrt{X^2+Y^2} = R$. But, $\mathbf{E}\left[R\right] = \displaystyle\int_0^{\varepsilon}\int_0^{\pi/3} n\, r\, G(r,\theta)^{n-1} f(r,\theta) r\, dr\, d\theta$. Integrand is critical at $r=0$, since for $r>0$ it converges to zero as $n\rightarrow\infty$. So we use the Taylor series expansion around $r=0$ as

$$\begin{aligned} f(r,\theta) &= f(0,\theta) + \frac{\partial f(0,\theta)}{\partial r} r + O\left(r^2\right), \\ G(r,\theta) &= G(0,\theta) + \frac{\partial G(0,\theta)}{\partial r} r + O\left(r^2\right) = 1 + \frac{\partial G(0,\theta)}{\partial r} r + O\left(r^2\right). \end{aligned}$$

Note that $\dfrac{\partial G(0,\theta)}{\partial r} < 0$, since area of $S_R(u,v)$ decreases as $r$ increases for fixed $\theta$. So let $r = w/n$, then

$$\begin{aligned} \mathbf{E}\left[R\right] &\sim \int_0^{n\,\varepsilon}\int_0^{\pi/3} n\,\frac{w}{n}\left(1 + \frac{\partial G(0,\theta)}{\partial r}\frac{w}{n} + O\left(n^{-2}\right)\right)^{n-1}\left(f(0,\theta) + \frac{\partial f(0,\theta)}{\partial r}\frac{w}{n} + O\left(n^{-2}\right)\right)\frac{w}{n^2}\, dw\, d\theta \\ &= \frac{1}{n^2}\int_0^{\infty}\int_0^{\pi/3} w^2 \exp\left(\frac{\partial G(0,\theta)}{\partial r} w\right) f(0,\theta) w\, dw\, d\theta = O\left(n^{-2}\right). \end{aligned}$$

Hence $P(X_{e_2} = X_{e_3} \neq \mathsf{y}_1) \rightarrow 0$ as $n\rightarrow\infty$. Then $P(X_{e_2} = X_{e_3}) \rightarrow 0$ as $n \rightarrow \infty$.

Likewise, it follows that $\lim_{n\rightarrow\infty} P(X_{e_1} = X_{e_2}) = \lim_{n\rightarrow\infty} P(X_{e_1} = X_{e_3}) = 0$. Hence $P(E_{c,2}(n)) \rightarrow 0$ as $n\rightarrow\infty$. Thus $P(E_{c,3}(n)) \rightarrow 1$ as $n\rightarrow\infty$. $\blacksquare$

Notice that Theorem 6.4 follows as a corollary from Theorem 6.6. For $r \geq 3/2$ and $M \in \mathbb{R}^2 \setminus \mathcal{Y}_3$, $\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right) \neq \emptyset$ a.s., since $\mathscr{R}_S\left(N_{PE}^r, M\right) \neq \emptyset$ and $\mathscr{R}_S\left(N_{PE}^r, M\right) \subseteq \Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right)$.

Now, for $n>1$, let $X_{e_i}(n) = x_{e_i} = (u_i, w_i)$ be given for $i \in \{1,2,3\}$, be the edge extrema in a given realization of $\mathcal{X}_n$. Then the functional form of $\Gamma_1$-region in $T_b$ is given by

$$\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right) = \bigcup_{i=1}^{3}\left[\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right) \cap R_M(\mathsf{y}_i)\right]$$

where

$$\begin{aligned} \Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right) \cap R_M(\mathsf{y}_1) &= \left\{(x,y) \in R_M(\mathsf{y}_1) : y \geq \frac{w_1}{r} - \frac{c_2\,(r\,x - u_1)}{(1-c_1)\,r}\right\}, \\ \Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right) \cap R_M(\mathsf{y}_2) &= \left\{(x,y) \in R_M(\mathsf{y}_1) : y \geq \frac{w_2}{r} - \frac{c_2\,(r\,(x-1)+1-u_2}{c_1\,r}\right\}, \\ \Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right) \cap R_M(\mathsf{y}_3) &= \left\{(x,y) \in R_M(\mathsf{y}_1) : y \leq \frac{w_3 - c_2\,(1-r)}{r}\right\}. \end{aligned}$$

See Figure 10 for $\Gamma_1\left(\mathcal{X}_n, N_{PE}^2, M\right)$ with $n \geq 3$ where $M$-vertex regions for $M = M_{CC}$ and $M = M_I$ with orthogonal projections are used. Note that only the edge extrema are shown in Figure 10 (right).

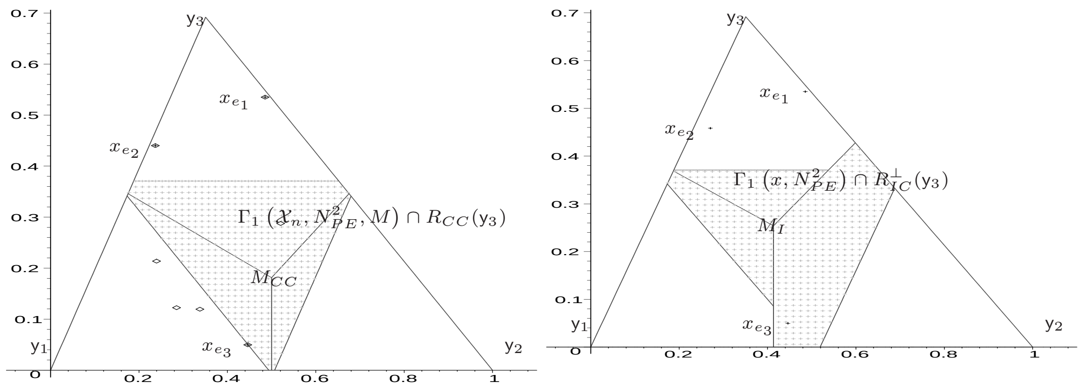


Figure 10: $\Gamma_1\left(\mathcal{X}_n, N_{PE}^2, M\right)$ for $M = M_{CC}$ (left) and $M = M_I$ (right) with three distinct edge extrema.

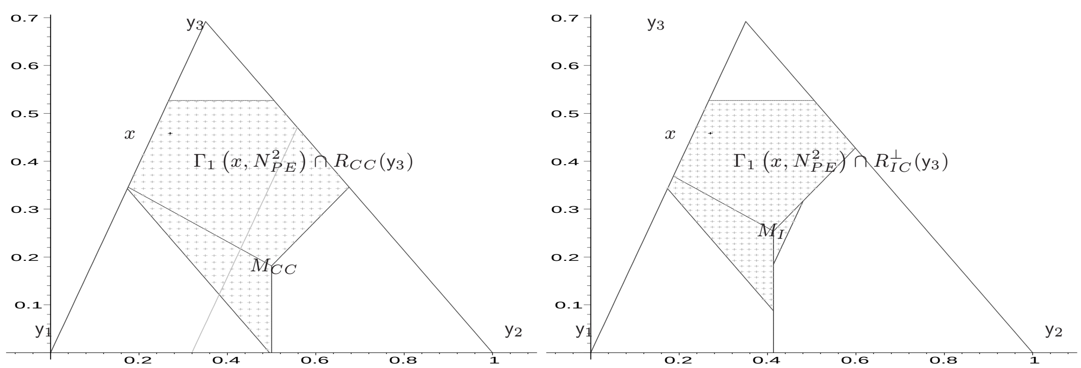


Figure 11: $\Gamma_1\left(x, N_{PE}^{\sqrt{2}}, M\right)$ with $x \in R_{CC}(\mathsf{y}_3)$ (left), $x \in R_{IC}^{\perp}(\mathsf{y}_3)$.

Note that, for $\mathcal{X}_n$ a random sample from $\mathcal{U}(T(\mathcal{Y}_3))$, $P(\eta_n(N_{PE}^r) = 3) \to 1$ as $n \to \infty$, since the edge extrema are distinct with probability 1 as $n \to \infty$. However, for $r < 3/2$, the region $\Gamma_1(x, N_{PE}^r, M) \cap R_M(\mathsf{y}_i)$ might be empty for some $i \in \{1, 2, 3\}$. Furthermore, if $M \in (\mathscr{T}^r)^o$ (see Equation (2) for $\mathscr{T}^r$) with $r < 3/2$, then $\Gamma_1(\mathcal{X}_n, N_{PE}^r, M)$ will be empty with probability 1 as $n \to \infty$. In such a case, there is no $\Gamma_1$-region to construct. But the definition of the $\eta$-value still works in the sense that $\Gamma_1(X_n, N_{PE}^r, M) = \emptyset = \bigcap_{x \in S_M} \Gamma_1(x, N_{PE}^r, M)$ (see Definition 4.4 for $S_M$) and $\Gamma_1(x, N_{PE}^r, M) \neq \emptyset$ for all $x \in \mathcal{X}_n$ since $x \in \Gamma_1(x, N_{PE}^r, M)$. To determine whether the $\Gamma_1$-region is empty or not, it suffices to check the intersection of the $\Gamma_1$-regions of the edge extrema. If $M \notin (\mathscr{T}^r)^o$, the $\Gamma_1$-region is guaranteed to be nonempty.

Note that $\eta_n(N_{PE}^{r_1}) \stackrel{d}{=} \eta_n(N_{PE}^{r_2})$ for all $(r_1, r_2) \in [1, \infty)^2$, where $\stackrel{d}{=}$ stands for "equality in distribution".

See Figure 12 for $\Gamma_1\left(\mathcal{X}_n, N_{PE}^{\sqrt{2}}, M\right)$ where $M$-vertex regions for $M = M_{CC}$ and $M = M_I$ with orthogonal projections are used.

*Remark* 6.7.

- For $r_1 < r_2$, $\Gamma_1(x, N_{PE}^{r_1}, M) \subseteq \Gamma_1(x, N_{PE}^{r_2}, M)$ for all $x \in T(\mathcal{Y}_3)$ with equality holding only when $x \in \mathcal{Y}_3$.
- For $r_1 < r_2$, $\Gamma_1(\mathcal{X}_n, N_{PE}^{r_1}, M) \subseteq \Gamma_1(\mathcal{X}_n, N_{PE}^{r_2}, M)$ with equality holding only when $\mathcal{X}_n \subseteq \mathcal{Y}_3$ or $\Gamma_1(\mathcal{X}_n, N_{PE}^r, M) = \emptyset$ for $r = r_1, r_2$.

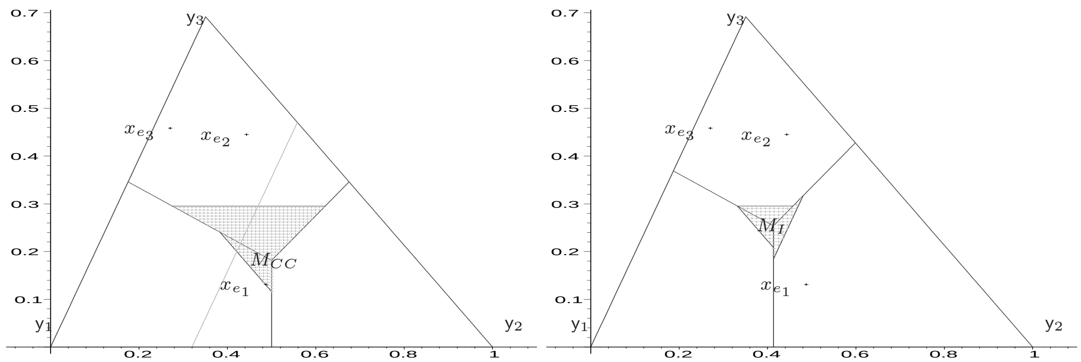


Figure 12: $\Gamma_1\left(\mathcal{X}_n, N_{PE}^{\sqrt{2}}, M\right)$ for $M = M_{CC}$ (left) and $M = M_I$ (right) with three distinct edge extrema.

- Suppose $X, Y$ are iid from a continuous distribution $F$ whose support is $\mathcal{S}(F) \subseteq T(\mathcal{Y}_3)$. Then for $r_1 < r_2$, $A\left(\Gamma_1\left(X, N_{PE}^{r_1}, M\right)\right) \leq^{ST} A\left(\Gamma_1\left(Y, N_{PE}^{r_2}, M\right)\right)$.
- Suppose $\mathcal{X}_n$ and $\mathcal{X}_n'$ are two random samples from a continuous distribution $F$ whose support is $\mathcal{S}(F) \subseteq T(\mathcal{Y}_3)$. Then for $r_1 < r_2$, $A(\Gamma_1(\mathcal{X}_n, N_{PE}^{r_1}, M)) \leq^{ST} A(\Gamma_1(\mathcal{X}_n', N_{PE}^{r_2}, M))$. $\square$

*Remark* 6.8. In $\mathbb{R}^d$ with $d > 2$, recall $\mathfrak{S}(\mathcal{Y}_m)$, the simplex based on $d+1$ points that do not lie on the same hyperplane. Furthermore, let $\varrho_i(r, x)$ be the hyperplane such that $\varrho_i(x) \cap \mathfrak{S}(\mathcal{Y}_m) \neq \emptyset$ and $r\, d(\mathsf{y}_i, \varrho_i(r, x)) = d(\mathsf{y}_i, \Upsilon(\mathsf{y}_i, x))$ for $i \in \{1, 2, \ldots, d+1$. Then

$$\Gamma_1\left(x, N_{PE}^r, M_C\right) \cap R_{CM}(\mathsf{y}_i) = \{z \in R_{CM}(\mathsf{y}_i) : d(\mathsf{y}_i, \Upsilon(\mathsf{y}_i, z)) \geq d(\mathsf{y}_i, \varrho_i(r, x)\} \text{ for } i \in \{1, 2, 3\}.$$

Hence $\Gamma_1\left(x, N_{PE}^r, M_C\right) = \bigcup_{i=1}^{d+1}(\Gamma_1\left(x, N_{PE}^r, M_C\right) \cap R_{CM}(\mathsf{y}_i))$. Furthermore, it is easy to see that $\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M_C\right) = \bigcap_{i=1}^{d+1}\Gamma_1\left(X_{\varphi_i}(n), N_{PE}^r, M_C\right)$, where $X_{\varphi_i}(n)$ is one of the closest points in $\mathcal{X}_n$ to face $\varphi_i$. $\square$

## 6.2 Central Similarity Proximity Maps

The other type of triangular proximity map we introduce is the central similarity proximity map. Furthermore, the relative density of the corresponding PCD will have mathematical tractability. Alas, the distribution of the domination number of the associated PCD is still an open problem.

For $\tau \in [0, 1]$, define $N_{CS}^\tau(\cdot, M) := N(\cdot, M; \tau, \mathcal{Y}_3)$ to be the *central similarity proximity map* with $M$-edge regions as follows; see also Figure 13 with $M = M_C$. For $x \in T(\mathcal{Y}_3) \setminus \mathcal{Y}_3$, let $e(x)$ be the edge in whose region $x$ falls; i.e., $x \in R_M(e(x))$. If $x$ falls on the boundary of two edge regions, we assign $e(x)$ arbitrarily. For $\tau \in (0, 1]$, the parametrized central similarity proximity region $N_{CS}^\tau(x, M)$ is defined to be the triangle $T_\tau(x)$ with the following properties:

(i) $T_\tau(x)$ has edges $e_i^\tau(x)$ parallel to $e_i$ for each $i \in \{1, 2, 3\}$, and for $x \in R_M(e(x))$, $d(x, e^\tau(x)) = \tau\, d(x, e(x))$ and $d(e^\tau(x), e(x)) \leq d(x, e(x))$ where $d(x, e(x))$ is the Euclidean (perpendicular) distance from $x$ to $e(x)$;

(ii) $T_\tau(x)$ has the same orientation as and is similar to $T(\mathcal{Y}_3)$;

(iii) $x$ is the same type of center of $T_\tau(x)$ as $M$ is of $T(\mathcal{Y}_3)$.

Note that (i) implies the parametrization of the proximity region, (ii) explains "similarity", and (iii) explains "central" in the name, *central similarity proximity map*. For $\tau = 0$, we define $N_{CS}^{\tau=0}(x, M) = \{x\}$ for all $x \in T(\mathcal{Y}_3)$. For $x \in \partial(T(\mathcal{Y}_3))$, we define $N_{CS}^\tau(x, M) = \{x\}$ for all $\tau \in [0, 1]$.

Notice that by definition $x \in N_{CS}^{\tau}(x, M)$ for all $x \in T(\mathcal{Y}_3)$. Furthermore, $\tau \leq 1$ implies that $N_{CS}^{\tau}(x, M) \subseteq T(\mathcal{Y}_3)$ for all $x \in T(\mathcal{Y}_3)$ and $M \in T(\mathcal{Y}_3)^o$. For all $x \in T(\mathcal{Y}_3)^o \cap R_M(e(x))$, the edges $e^{\tau}(x)$ and $e(x)$ are coincident iff $\tau = 1$.

Notice that $X_i \overset{iid}{\sim} F$, with the additional assumption that the non-degenerate two-dimensional probability density function $f$ exists with support $\mathcal{S}(F) \subseteq T(\mathcal{Y}_3)$, implies that the special case in the construction of $N_{CS}^{\tau}(\cdot)$ — $X$ falls on the boundary of two edge regions — occurs with probability zero. Note that for such an $F$, $N_{CS}^{\tau}(X, M)$ is a triangle a.s.

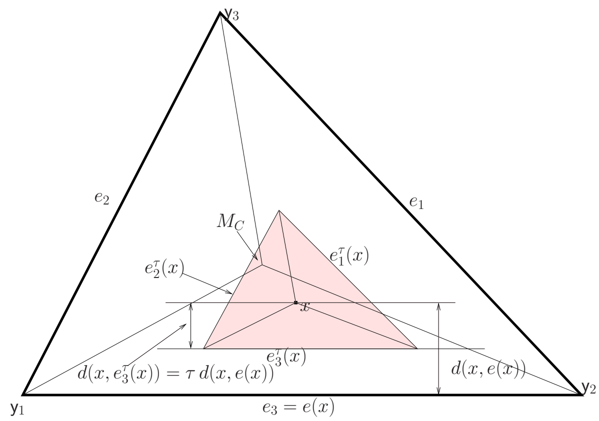


Figure 13: Construction of central similarity proximity region, $N_{CS}^{\tau=1/2}(x, M_C)$ (shaded region).

Central similarity proximity maps are defined with $M$-edge regions for $M \in T(\mathcal{Y}_3)^o$. In general, for central similarity proximity regions with $M$-edge regions, the similarity ratio of $N_{CS}^{\tau}(x, M)$ to $T(\mathcal{Y}_3)$ is $d(x, e^{\tau}(x))/d(M, e(x))$. See Figure 13 for $N_{CS}^{\tau=1/2}(x, M_C)$ with $e = e_3$. The functional form of $N_{CS}^{\tau}(x, M_C)$ is provided in Ceyhan (2009).

### 6.2.1 Extension of $N_{CS}^{\tau}$ to Higher Dimensions

The extension of $N_{CS}^{\tau}$ to $\mathbb{R}^d$ for $d > 2$ is straightforward. the extension for $M = M_C$ is described, the extension for general $M$ is similar. Let $\mathcal{Y}_{d+1} = \{\mathsf{y}_1, \mathsf{y}_2, \ldots, \mathsf{y}_{d+1}\}$ be $d+1$ points that do not lie on the same $(d-1)$-dimensional hyperplane. Denote the simplex formed by these $d+1$ points as $\mathfrak{S}(\mathcal{Y}_{d+1})$. For $\tau \in (0, 1]$, define the central similarity proximity map as follows. Let $\varphi_i$ be the face opposite vertex $\mathsf{y}_i$ for $i \in \{1, 2, \ldots, (d+1)\}$, and "face regions" $R_{CM}(\varphi_1), R_{CM}(\varphi_2), \ldots, R_{CM}(\varphi_{d+1})$ partition $\mathfrak{S}(\mathcal{Y}_{d+1})$ into $d+1$ regions, namely the $d+1$ polytopes with vertices being the center of mass together with $d$ vertices chosen from $d+1$ vertices. For $x \in \mathfrak{S}(\mathcal{Y}_{d+1}) \setminus \mathcal{Y}_{d+1}$, let $\varphi(x)$ be the face in whose region $x$ falls; $x \in R(\varphi(x))$. If $x$ falls on the boundary of two face regions, $\varphi(x)$ is assigned arbitrarily. For $\tau \in (0, 1]$, the central similarity proximity region $N_{CS}^{\tau}(x, M_C) = \mathfrak{S}_{\tau}(x)$ is defined to be the simplex $\mathfrak{S}_{\tau}(x)$ with the following properties:

(i) $\mathfrak{S}_{\tau}(x)$ has faces $\varphi_i^{\tau}(x)$ parallel to $\varphi_i(x)$ for $i \in \{1, 2, \ldots, (d+1)\}$, and for $x \in R_{CM}(\varphi(x))$, $\tau\, d(x, \varphi(x)) = d(\varphi^{\tau}(x), x)$ where $d(x, \varphi(x))$ is the Euclidean (perpendicular) distance from $x$ to $\varphi(x)$;

(ii) $\mathfrak{S}_{\tau}(x)$ has the same orientation as and similar to $\mathfrak{S}(\mathcal{Y}_{d+1})$;

(iii) $x$ is the center of mass of $\mathfrak{S}_{\tau}(x)$, as $M_C$ is of $\mathfrak{S}(\mathcal{Y}_{d+1})$. Note that $\tau > 1$ implies that $x \in N_{CS}^{\tau}(x)$.

### 6.2.2 $\Gamma_1$-Regions for Central Similarity Proximity Maps

For $N_{CS}^{\tau}$, the $\Gamma_1$-region is constructed as follows. Let $e_i^{\tau}(x)$ be the edge of $T_{\tau}(x)$ parallel to edge $e_i$ for $i \in \{1,2,3\}$. Now, suppose $u \in R_M(e_3)$ and let $\zeta_i(\tau,x)$ for $i \in \{1,2,\ldots,7\}$ be the lines such that

$$\begin{aligned}
&v \in \zeta_1(\tau,u) \cap R_M(e_3) \Longrightarrow u \in e_1^{\tau}(v), && v \in \zeta_5(\tau,u) \cap R_M(e_2) \Longrightarrow u \in e_3^{\tau}(v),\\
&v \in \zeta_2(\tau,u) \cap R_M(e_3) \Longrightarrow u \in e_2^{\tau}(v), && v \in \zeta_6(\tau,u) \cap R_M(e_2) \Longrightarrow u \in e_2^{\tau}(v),\\
&v \in \zeta_3(\tau,u) \cap R_M(e_1) \Longrightarrow u \in e_2^{\tau}(v), && v \in \zeta_7(\tau,u) \cap R_M(e_3) \Longrightarrow u \in e_3^{\tau}(v).\\
&v \in \zeta_4(\tau,u) \cap R_M(e_1) \Longrightarrow u \in e_3^{\tau}(v), &&
\end{aligned} \tag{3}$$

Then $\Gamma_1\left(x, N_{CS}^{\tau}, M\right)$ is the region bounded by these lines. See also Figure 14. $\Gamma_1\left(x, N_{CS}^{\tau}, M\right)$ for $x \in R_M(e_i)$ for $i \in \{1,2\}$ can be described similarly.

Notice that $\tau > 0$ implies that $x \in \Gamma_1\left(x, N_{CS}^{\tau}, M\right)$. Furthermore, $\Gamma_1\left(x, N_{CS}^{\tau}, M\right) = \{x\}$ iff (i) $\tau = 0$ or (ii) $x \in \partial(T(\mathcal{Y}_3))$.

The $\Gamma_1$-region $\Gamma_1\left(x, N_{CS}^{\tau}, M\right)$ is a convex $k$-gon with $3 \le k \le 7$ vertices. In particular, for $\tau = 1$, $\Gamma_1\left(x, N_{CS}^{\tau=1}, M\right)$ is a convex hexagon. See Figure 16 (left).

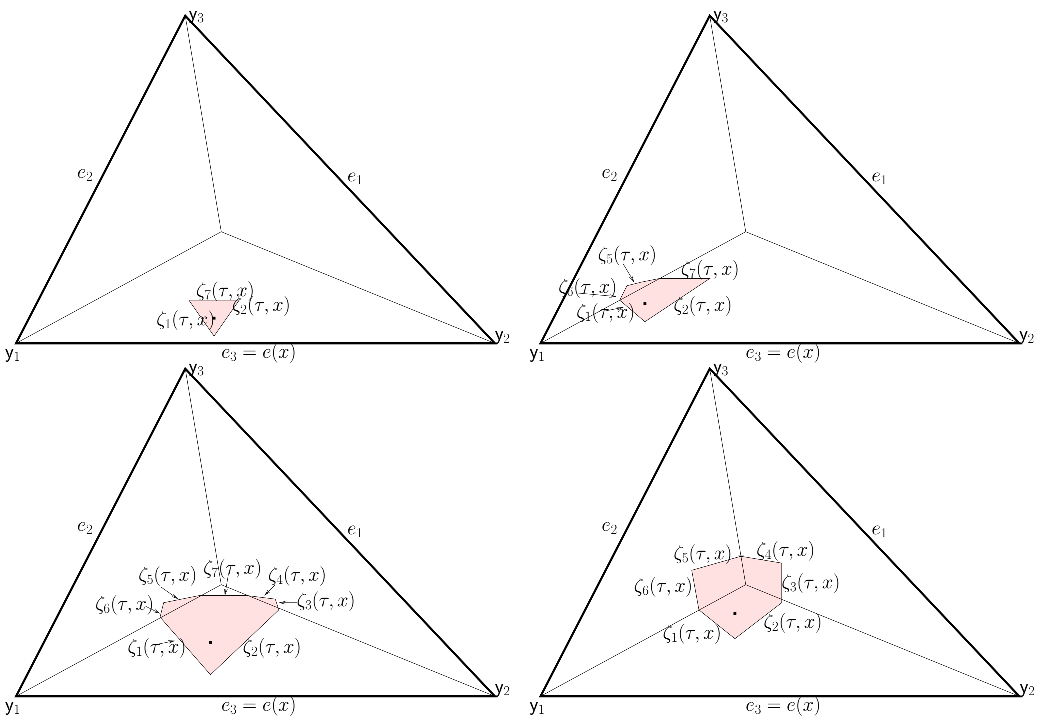


Figure 14: Examples of the four types of the $\Gamma_1$-region, $\Gamma_1\left(x, N_{CS}^{\tau=1/2}, M_C\right)$ with four distinct $x \in R_{CM}(e_3)$ (shaded regions).

Then the functional form of $\zeta_i(x)$ in Equation (3) that define the boundary of $\Gamma_1\left(x = (x_0, y_0), N_{CS}^{\tau}\right)$ with

$M$-central proximity regions in the basic triangle $T_b$ for $x \in R_M(e_3)$ is

$$\zeta_1(x) = \frac{m_2\,(c_2\,(x-x_0)+y_0\,c_1)}{m_2\,c_1\,(1-\tau)+c_2\,\tau\,m_1},\ \ \zeta_2(x) = \frac{m_2\,(c_2\,(x_0-x)+y_0\,(1-c_1))}{m_2\,(1-c_1)\,(1-\tau)+c_2\,\tau\,(1-m_1)},$$

$$\zeta_3(x) = \frac{m_2\,c_2\,(x-1)+y_0\,(m_2\,(1-c_1)-\,c_2\,(1-m_2))}{c_1\,(1-c_1)\,(1-\tau)},$$

$$\begin{aligned}\zeta_4(x) = \Big[&c_1\,(m_2\,(1-c_1)+c_2\,(1-m_1))\,y_0+c_2\,(m_2\,(1-c_1)+c_2\,(1-m_1))\,x_0+c_2\,(\tau\,(m_2\,c_1-m_1\,c_2)\\ &+c_2\,(1-m_1)+m_2\,(1-c_1))\,x+c_2\,\tau\,(m_1\,c_2-m_2\,c_1)\Big]\Big/\Big[c_2\,(\tau-c_1\,(1-\tau))\,m_1\\ &+c_1\,(1-c_1)\,(1-\tau)\,m_2+c_1\,c_2\Big],\end{aligned}$$

$$\zeta_5(x) = \frac{\tau\,m_2\,c_2\,x-y_0\,(c_1\,m_2-c_2\,m_1)}{c_2\,m_1-c_1\,m_2\,(1-\tau)},\ \ \zeta_7(x) = \frac{y_0}{1-\tau}$$

$$\begin{aligned}\zeta_6(x) = \Big[&(1-c_1)\,c_1\,(m_2\,(1-c_1)-c_2\,(1-m_1))\,y_0+c_1\,c_2\,(m_2\,(1-c_1)-c_2\,(1-m_1))\,x_0+(c_2^2\,c_1\,(1-\tau)\,m_1\\ &+c_2\,\tau\,(1-c_1)+(1-\tau)\,(c_1\,c_2-c_1^2\,c_2)\,m_2-c_2^2\,\tau\,(1-c_1)-c_2^2\,c_1\,(1-\tau))\,x+c_2^2\,\tau\,(m_1-c_1)\Big]\Big/\Big[(1-c_1)\\ &(c_2\,(c_2\,(1-\tau)-\tau)\,m_1+c_1\,(1-c_1)\,(1-\tau)\,m_2+c_1\,c_2\,(1-2\,\tau))\Big]\end{aligned}$$

**Proposition 6.9.** *Let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(T(\mathcal{Y}_3))$. For central similarity proximity maps with $M$-edge regions (by definition $M \in T(\mathcal{Y}_3)^o$) and $\tau > 0$, $\eta_n(N_{CS}^\tau) \le 3$ with equality holding with positive probability for $n \ge 3$.*

**Proof:** Let $M \in T(\mathcal{Y}_3)^o$ and $\tau > 0$. Then given $\mathcal{X}_n$, for $i \in \{1,2,3\}$, we have

$$\begin{aligned}\Gamma_1(\mathcal{X}_n, N_{CS}^\tau, M)) \cap R_M(e_i) &= \bigcap_{i=1}^{n} [\Gamma_1(X_i, N_{CS}^\tau, M) \cap R_M(e_i)]\\ &= \Gamma_1\left(X_{e_j}(n), N_{CS}^\tau, M\right) \cap \Gamma_1(X_{e_k}(n), N_{CS}^\tau, M) \cap R_M(e_i)\end{aligned}$$

where $j,k \ne i$, since for $x \in R_M(e_i)$, if $\{X_{e_k}(n), X_{e_l}(n)\} \subset N_{CS}^\tau(x,M)$, then $\mathcal{X}_n \subset N_{CS}^\tau(x,M)$. Hence for each edge we need the edge extrema with respect to the other edges, then the minimum active set is $S_M = \{X_{e_1}(n), X_{e_2}(n), X_{e_3}(n)\}$, hence $\eta_n(N_{CS}^\tau) \le 3$. Furthermore, for the random sample $\mathcal{X}_n$, $X_e(n)$ is unique for each edge $e$ with probability 1 and there are three distinct edge extrema with positive probability (see Theorem 6.4). Hence $P(\eta_n(N_{CS}^\tau) = 3) > 0$ for $n \ge 3$. ■

Note that for $\tau > 0$ and $\mathcal{X}_n$ a random sample from $\mathcal{U}(T(\mathcal{Y}_3))$, $P(\eta(\mathcal{X}_n, N_{CS}^\tau) = 3) \to 1$ as $n \to \infty$, since the edge extrema are distinct with probability 1 as $n \to \infty$. For $\tau = 1$, the $\Gamma_1$-region can be determined by the edge extrema, in particular $\Gamma_1\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M\right) \cap R_M(\mathsf{y}_i)$ can be determined by the edge extrema $X_{e_j}$ and $X_{e_k}$ for $i,j,k$ are all distinct. See Figure 15 for an example. Here $\Gamma_1\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M\right) \ne \emptyset$ for all $n$, because by construction $M \in \Gamma_1\left(\mathcal{X}_n, N_{CS}^{\tau=1}, M\right)$ since $N_{CS}^{\tau=1}(M) = T(\mathcal{Y}_3)$. Furthermore, $\eta_n(N_{CS}^\tau) \stackrel{d}{=} \eta_n(N_{PE}^r)$ for all $(r,\tau) \in [1,\infty) \times (0,1]$.

**Proposition 6.10.** *The $\Gamma_1$-region $\Gamma_1\left(B, N_{CS}^{\tau=1}, M\right)$ is a convex hexagon for any set $B$ of size $n$ in $T(\mathcal{Y}_3)^o$.*

**Proof:** This follows from the fact that each $\zeta_i(x)$ for $i \in \{1,2,\ldots,6\}$ is parallel to a line joining $\mathsf{y}_i$ to $M$ for $\tau = 1$ ($\zeta_7(x)$ is not used in construction of $\Gamma_1\left(B, N_{CS}^{\tau=1}, M\right)$). See Figure 16 (right) for an example. ■

With $M = M_C$, the functional forms of the lines that determine the boundary of $\Gamma_1(x = (x_0, y_0), N_{CS}^\tau)$

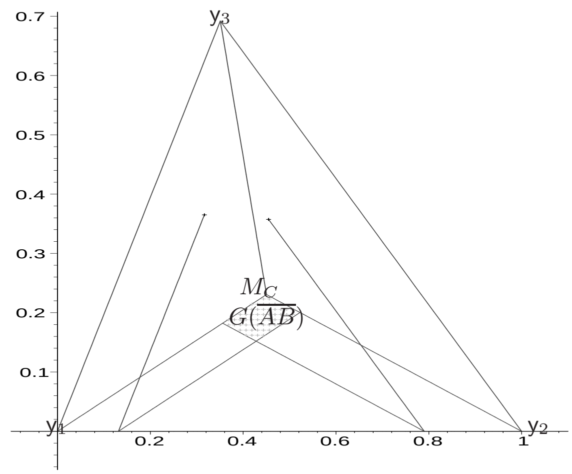


Figure 15: $G(\overline{AB})$ region for $x_1$ and $\Gamma_1$-region for $n \geq 1$

in Equation (3) for $T_b$ with $x \in R_{CM}(e_3)$ are given by

$$\begin{aligned}
&\zeta_1(x) = \frac{y_0\, c_1 + c_2\,(x - x_0)}{c_1 + \tau},\ \zeta_2(x) = \frac{c_2\,(x_0 - x) + y_0\,(1 - c_1)}{\tau + 1 - c_1},\\
&\zeta_3(x) = \frac{c_2\,\tau\,(1 - x) + y_0}{1 + \tau\,(1 - c_1)}, \quad \zeta_4(x) = \frac{\tau\, c_2\,(1 - x) + c_2\,(x_0 - x) - y_0\, c_1}{\tau\,(1 - c_1) - c_1},\\
&\zeta_5(x) = \frac{c_2\, x + y_0}{1 + \tau\, c_1}, \qquad \zeta_7(x) = \frac{y_0}{1 - \tau},\\
&\zeta_6(x) = \frac{c_1\,(1 - c_1)\, y_0 + c_1\, c_2\, x_0 - c_2\,(2\,\tau\,(1 - c_1) + c_1\,(1 - \tau))\, x + \tau\, c_2\,(1 - 2\, c_1)}{c_1\,(1 - c_1)\,(1 - \tau)}.
\end{aligned}$$

For $n > 1$ with $M = M_C$, let $X_{e_i}(n) = x_{e_i} = (u_i, w_i)$ be given for $i \in \{1, 2, 3\}$. Then the functional forms of the lines that determine the boundary of $\Gamma_1(\mathcal{X}_n, N^{\tau}_{CS}, M_C)$ are given by

$$\begin{aligned}
&\zeta_1(x) = \frac{w_2\, c_1 + c_2\,(x - u_2)}{c_1 + \tau}, \qquad \zeta_3(x) = \frac{c_2\,\tau\,(1 - x) + y_0}{1 + \tau\,(1 - c_1)},\\
&\zeta_2(x) = \frac{c_2\,(u_1 - x) + w_1\,(1 - c_1)}{\tau + 1 - c_1}, \qquad \zeta_5(x) = \frac{c_2\, x + w_3}{1 + \tau\, c_1},\\
&\zeta_4(x) = \frac{\tau\, c_2\,(1 - x) + c_2\,(u_3 - x) - w_3\, c_1}{\tau\,(1 - c_1) - c_1}, \quad \zeta_7(x) = \frac{w_3}{1 - \tau},\\
&\zeta_6(x) = \frac{c_1\,(1 - c_1)\, w_1 + c_1\, c_2\, u_1 - c_2\,(2\,\tau\,(1 - c_1) + c_1\,(1 - \tau))\, x + \tau\, c_2\,(1 - 2\, c_1)}{c_1\,(1 - c_1)\,(1 - \tau)}.
\end{aligned}$$

See Figure 16 for $\Gamma_1\left(x, N^{\tau=1}_{CS}, M_C\right)$ and $\Gamma_1\left(\mathcal{X}_n, N^{\tau=1}_{CS}, M_C\right)$.

*Remark* 6.11.

- For $\tau_1 < \tau_2$, $\Gamma_1(x, N^{\tau_1}_{CS}, M) \subseteq \Gamma_1(x, N^{\tau_2}_{CS}, M)$ for all $x \in T(\mathcal{Y}_3)$ with equality holding only when $x \in \partial(T(\mathcal{Y}_3))$.
- Let $\mathcal{X}_n$ be a set of $n$ points in $T(\mathcal{Y}_3)$. Then for $\tau_1 < \tau_2$, $\Gamma_1(\mathcal{X}_n, N^{\tau_1}_{CS}, M) \subseteq \Gamma_1(\mathcal{X}_n, N^{\tau_2}_{CS}, M)$ with equality holding only when $\mathcal{X}_n \subset \partial(T(\mathcal{Y}_3))$.
- Suppose $X, Y$ are iid from a continuous distribution $F$ whose support is $\mathcal{S}(F) \subseteq T(\mathcal{Y}_3)$. Then for $\tau_1 < \tau_2$, $A(\Gamma_1(X, N^{\tau_1}_{CS}, M)) \leq^{ST} A(\Gamma_1(Y, N^{\tau_2}_{CS}, M))$.

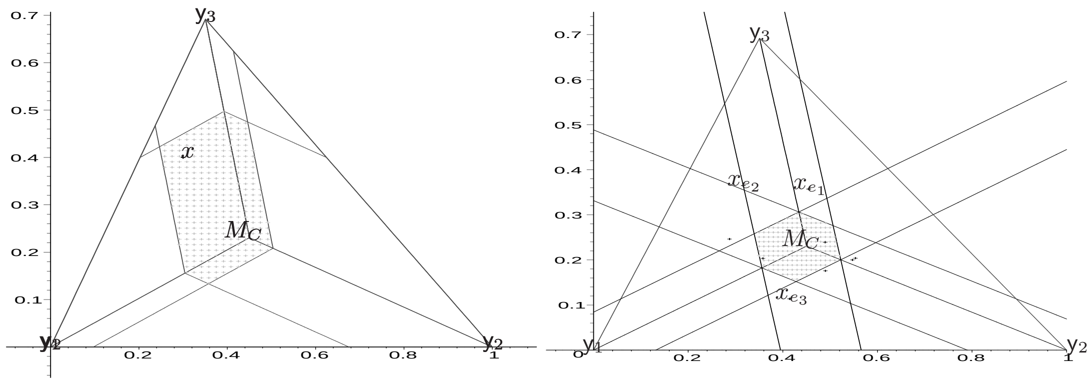


Figure 16: The $\Gamma_1$-region $\Gamma_1\left(x, N_{CS}^{\tau}, M_C\right)$ with $x \in R_M(e_2)$ (left) and $\Gamma_1\left(X_n, N_{CS}^{\tau}, M_C\right)$ with $n > 1$ (right).

- Suppose $\mathcal{X}_n$ and $\mathcal{X}'_n$ are two random samples from a continuous distribution $F$ whose support is $\mathcal{S}(F) \subseteq T\left(\mathcal{Y}_3\right)$. Then for $\tau_1 < \tau_2$, $A(\Gamma_1\left(\mathcal{X}_n, N_{CS}^{\tau_1}, M\right)) \leq^{ST} A(\Gamma_1\left(\mathcal{X}'_n, N_{CS}^{\tau_2}, M\right))$. $\square$

# 7 Investigation of the Proximity Regions and the Associated PCDs Using the Auxiliary Tools

## 7.1 Characterization of Proximity Maps Using $\eta$-Values

By definition, it is trivial to show that that the minimum number of points to describe the $\Gamma_1$-region, $\eta_n(N) \leq n$ for any proximity map. We have improved the upper bound for $N_{PE}^r$ and $N_{CS}^{\tau}$: $\eta(\mathcal{X}_n, N_{PE}^r) \leq 3$ and $\eta(\mathcal{X}_n, N_{CS}^{\tau}) \leq 3$. However, finding such an improvement does not hold for $\eta_n(N_{AS})$; that is, finding a $k < n$ such that $\eta_n(N_{AS}) \leq k$ for all $\mathcal{X}_n$ is still an open problem.

Below we state a condition for $N(\cdot, M)$ defined with $M$-vertex regions to have $\eta_n(N) \leq 3$ for $\mathcal{X}_n$ with support in $T\left(\mathcal{Y}_3\right)$.

**Theorem 7.1.** *Suppose $N(\cdot, M)$ is a proximity region defined with $M$-vertex regions and $B$ is a set of $n$ distinct points in $T\left(\mathcal{Y}_3\right)$. Then $\eta(B, N) \leq 3$ if*

*(i) for each $\mathsf{y}_i \in \mathcal{Y}_3$ there exists a point $x(\mathsf{y}_i) \in B$ (i.e., related to $\mathsf{y}_i$) such that $\Gamma_1(B, N) \cap R_M(\mathsf{y}_i) = \Gamma_1\left(x(\mathsf{y}_i), N\right) \cap R_M(\mathsf{y}_i)$,*

*or*

*(ii) there exist points $x(\mathsf{y}_j)$, $x(\mathsf{y}_k) \in B$ such that $\Gamma_1(B, N) \cap R_M(\mathsf{y}_i) = \Gamma_1\left(x(\mathsf{y}_j), N\right) \cap \Gamma_1\left(x(\mathsf{y}_k), N\right) \cap R_M(\mathsf{y}_i)$ for $j, k \neq i$ with $i \in \{1, 2, 3\}$ and $(j, k) \in \{(1, 2), (1, 3), (2, 3)\}$.*

**Proof:** Let $B = \{x_1, x_2, \ldots, x_n\} \subset T\left(\mathcal{Y}_3\right)$.

(i) Suppose there exists a point $x(\mathsf{y}_i) \in B$ such that $\Gamma_1(B, N) \cap R_M(\mathsf{y}_i) = \Gamma_1\left(x(\mathsf{y}_i), N\right) \cap R_M(\mathsf{y}_i)$ for each $i \in \{1, 2, 3\}$. Then

$$\Gamma_1(B, N) \cap R_M(\mathsf{y}_i) = \Gamma_1\left(x(\mathsf{y}_i), N\right) \cap R_M(\mathsf{y}_i) =$$
$$\bigcap_{i=1}^{n} [\Gamma_1\left(x_i, N\right) \cap R_M(\mathsf{y}_i)] = \bigcap_{j=1}^{3} [\Gamma_1\left(x(\mathsf{y}_j), N\right) \cap R_M(\mathsf{y}_i)] = \left[\bigcap_{j=1}^{3} \Gamma_1\left(x(\mathsf{y}_j), N\right)\right] \cap R_M(\mathsf{y}_i)$$

and

$$\Gamma_1(B,N) = \bigcup_{i=1}^{3}[\Gamma_1(B,N) \cap R_M(\mathsf{y}_i)] = \bigcup_{i=1}^{3}\left(\left[\bigcap_{j=1}^{3}\Gamma_1\left(x(\mathsf{y}_j),N\right)\right] \cap R_M(\mathsf{y}_i)\right).$$

Then, we get

$$\Gamma_1(B,N) = \bigcap_{i=1}^{3}\Gamma_1\left(x(\mathsf{y}_i),N\right).$$

Hence, the minimum active set $S_M \subseteq \{x(\mathsf{y}_1), x(\mathsf{y}_2), x(\mathsf{y}_3)\}$, which implies $\eta(B,N) \leq 3$. The $\eta$-value $\eta(B,N) < 3$ will hold if $x(\mathsf{y}_i)$ are not all distinct.

(ii) Suppose there exist points $x(\mathsf{y}_j)$ and $x(\mathsf{y}_k)$ such that $\Gamma_1\left(B,N\right) \cap R_M(\mathsf{y}_i) = \Gamma_1\left(x(\mathsf{y}_j),N\right) \cap \Gamma_1\left(x(\mathsf{y}_k),N\right) \cap R_M(\mathsf{y}_i)$ for $j,k \neq i$. Then

$$\Gamma_1\left(B,N\right) \cap R_M(\mathsf{y}_i) = \Gamma_1\left(x(\mathsf{y}_j),N\right) \cap \Gamma_1\left(x(\mathsf{y}_k),N\right) \cap R_M(\mathsf{y}_i) =$$
$$\bigcap_{i=1}^{n}[\Gamma_1\left(x_i,N\right) \cap R_M(\mathsf{y}_i)] = \bigcap_{q=1}^{3}[\Gamma_1\left(x(\mathsf{y}_q),N\right) \cap R_M(\mathsf{y}_i)] = \left[\bigcap_{q=1}^{3}\Gamma_1\left(x(\mathsf{y}_q),N\right)\right] \cap R_M(\mathsf{y}_i)$$

and

$$\Gamma_1(B,N) = \bigcup_{i=1}^{3}[\Gamma_1(B,N) \cap R_M(\mathsf{y}_i)] = \bigcup_{i=1}^{3}\left(\left[\bigcap_{q=1}^{3}\Gamma_1\left(x(\mathsf{y}_q),N\right)\right] \cap R_M(\mathsf{y}_i)\right).$$

Then, we get $\Gamma_1(B,N) = \bigcap_{i=1}^{3}\Gamma_1\left(x(\mathsf{y}_i),N\right)$. Hence, the minimum active set $S_M \subseteq \{x(\mathsf{y}_1), x(\mathsf{y}_2), x(\mathsf{y}_3)\}$ which implies $\eta(B,N) \leq 3$. ■

Notice that $N_{PE}^{r}$ satisfies condition (i) Theorem 7.1.

Below we state some conditions for $N(\cdot,M)$ defined with $M$-edge regions to have $\eta$-value less than equal to 3.

**Theorem 7.2.** *Suppose $N(\cdot,M)$ is a proximity region defined with $M$-edge regions and $B$ is set of $n$ distinct points in $T(\mathcal{Y}_3)$. Then $\eta(B,N) \leq 3$ if*

*(i) for each $e_i \in \{e_1,e_2,e_3\}$, there exists a point $x(e_i) \in B$ such that $\Gamma_1(B,N) \cap R_M(e_i) = \Gamma_1\left(x(e_i),N\right) \cap R_M(e_i)$,*

*or*

*(ii) there exist points $x(e_j)$, $x(e_k) \in B$ such that $\Gamma_1(B,N) \cap R_M(e_i) = \Gamma_1\left(x(e_j),N\right) \cap \Gamma_1\left(x(e_k),N\right) \cap R_M(e_i)$ for $j,k \neq i$ with $i \in \{1,2,3\}$ and $(j,k) \in \{(1,2),(1,3),(2,3)\}$.*

**Proof:** Let $B = \{x_1,x_2,\ldots,x_n\} \subset T(\mathcal{Y}_3)$.

(i) Suppose there exists a point $x(e_i) \in B$ such that $\Gamma_1(B,N) \cap R_M(e_i) = \Gamma_1\left(x(e_i),N\right) \cap R_M(e_i)$ for each $i \in \{1,2,3\}$. Then

$$\Gamma_1(B,N) \cap R_M(e_i) = \Gamma_1\left(x(e_i),N\right) \cap R_M(e_i) =$$
$$\bigcap_{i=1}^{n}[\Gamma_1\left(x_i,N\right) \cap R_M(e_i)] = \bigcap_{j=1}^{3}[\Gamma_1\left(x(e_j),N\right) \cap R_M(e_i)] = \left[\bigcap_{j=1}^{3}\Gamma_1\left(x(e_j),N\right)\right] \cap R_M(e_i)$$

and

$$\Gamma_1(B,N) = \bigcup_{i=1}^{3}[\Gamma_1(B,N) \cap R_M(e_i)] = \bigcup_{i=1}^{3}\left(\left[\bigcap_{j=1}^{3}\Gamma_1\left(x(e_j),N\right)\right] \cap R_M(e_i)\right).$$

Then, we get

$$\Gamma_1(\mathcal{X}_n, N) = \bigcap_{j=1}^{3} \Gamma_1(x(e_j), N).$$

Hence, the minimum active set $S_M \subseteq \{x(e_1), x(e_2), x(e_3)\}$ which implies $\eta_n(N) \leq 3$.

(ii) Suppose there exist points $x(e_j)$ and $x(e_k)$ such that $\Gamma_1(\mathcal{X}_n, N) \cap R_M(e_i) = \Gamma_1(x(e_j), N) \cap \Gamma_1(x(e_k), N) \cap R_M(e_i)$ for $j, k \neq i$. Then

$$\Gamma_1(B, N) \cap R_M(e_i) = \Gamma_1(x(e_j), N) \cap \Gamma_1(x(e_k), N) \cap R_M(e_i) =$$
$$\bigcap_{i=1}^{n} [\Gamma_1(x_i, N) \cap R_M(e_i)] = \bigcap_{q=1}^{3} [\Gamma_1(x(e_q), N) \cap R_M(e_i)] = \left[\bigcap_{q=1}^{3} \Gamma_1(x(e_q), N)\right] \cap R_M(e_i)$$

and

$$\Gamma_1(B, N) = \bigcup_{i=1}^{3} [\Gamma_1(B, N) \cap R_M(e_i)] = \bigcup_{i=1}^{3} \left(\left[\bigcap_{q=1}^{3} \Gamma_1(x(e_q), N)\right] \cap R_M(e_i)\right).$$

Then, we get $\Gamma_1(B, N) = \bigcap_{i=1}^{3} \Gamma_1(x(e_i), N)$. Hence, the minimum active set $S_M \subseteq \{x(e_1), x(e_2), x(e_3)\}$ which implies $\eta(B, N) \leq 3$. ■

Notice that $N_{CS}^{\tau}$ satisfies condition (ii) in Theorem 7.2.

## 7.2 The Behavior of $\Gamma_1(\mathcal{X}_n, N)$ for the Proximity Maps in $T(\mathcal{Y}_3)$

In Section 4, we have investigated the behavior of $\Gamma_1(\mathcal{X}_n, N)$ for general proximity maps in $\Omega$. The assertions made about $\Gamma_1$-regions will be stronger for the proximity regions we have defined, i.e., for $N_{AS}$, $N_{PE}^r$, and $N_{CS}^{\tau}$, compared to the general assertions in Section 4. One property enjoyed by these proximity maps is that the region $N(x)$ gets larger as $x$ moves along a line from $\partial(T(\mathcal{Y}_3))$ to $\mathscr{R}_S(N)$ in a region with positive $\mathbb{R}^2$-Lebesgue measure. So the modifications of the assertions in Section 4 also hold for $\{N_S, N_{AS}, N_{PE}^r, N_{CS}^{\tau}\}$. In particular, we have a stronger result than the one in Proposition 4.1 in the sense that, $\mathscr{R}_S(N)$ is a proper subset of $\Gamma_1(\mathcal{X}_n, N)$ as shown below.

**Proposition 7.3.** *For each type of proximity map $N \in \{N_S, N_{AS}, N_{PE}^r, N_{CS}^{\tau}\}$ and any random sample $\mathcal{X}_n = \{X_1, X_2, \ldots, X_n\}$ from a continuous distribution $F$ on $T(\mathcal{Y}_3)$, if $\mathscr{R}_S(N) \neq \emptyset$, then $\mathscr{R}_S(N) \subsetneq \Gamma_1(\mathcal{X}_n, N)$ a.s. for each $n < \infty$.*

**Proof:** We have shown that $\mathscr{R}_S(N) \subseteq \Gamma_1(\mathcal{X}_n, N)$ (see Proposition 4.1). Moreover, $\mathscr{R}_S(N) \neq \emptyset$ for $N_{AS}$, $N_{PE}^r$ with $r > 3/2$, and $N_{CS}^{\tau=1}$. For these proximity regions, $\mathscr{R}_S(N) = \Gamma_1(\mathcal{X}_n, N)$ with probability 0 for each finite $n$ since

(i) for $N \in \{N_S, N_{AS}\}$, $\Gamma_1(\mathcal{X}_n, N) = \mathscr{R}_S(N)$ iff $X_{\mathsf{y}}(n) = \mathsf{y}$ for each $\mathsf{y} \in \mathcal{Y}_3$ which happens with probability 0,

(ii) for $N \in \left\{N_{CS}^{\tau=1}, N_{PE}^{r>3/2}\right\}$, $\Gamma_1(\mathcal{X}_n, N, M) = \mathscr{R}_S(N, M)$ iff $X_{e_i}(n) \in e_i$ for each $i \in \{1, 2, 3\}$ which happens with probability 0.

Furthermore, for $N_{PE}^r$ with $r < 3/2$, $\mathscr{R}_S(N_{PE}^r, M) \neq \emptyset$ iff $M \notin (\mathscr{T}^r)^o$ (see Equation (2) for $\mathscr{T}^r$), say $M = (m_x, m_y)$ is such that $d(\zeta(m_x, \mathsf{y}_2), \mathsf{y}_2) < d(\mathsf{y}_2, e_2)/r$. Then $\Gamma_1(\mathcal{X}_n, N_{PE}^r, M) = \mathscr{R}_S(N_{PE}^r, M)$ iff $X_{e_2}(n) \in e_2$ which happens with probability 0. Similarly the same result also holds for edges $e_1$ and $e_3$. ■

Note that, if $\mathscr{R}_S(N) = \emptyset$ and $\mathcal{X}_n$ is a random sample from a continuous distribution on $T(\mathcal{Y}_3)$, then $\Gamma_1(\mathcal{X}_n, N) = \emptyset$ a.s. as $n \to \infty$. In particular, this holds for $N_{PE}^r(\cdot, M)$ with $r < 3/2$ and $M \in (\mathscr{T}^r)^o$. Lemma

4.6 holds as stated. In Lemma 4.6, we have shown that $\Gamma_1(\mathcal{X}(n), N)$ is non-increasing. Furthermore, for the proximity regions $\{N_S, N_{AS}, N_{PE}^r, N_{CS}^\tau\}$, we can state that $\Gamma_1(\mathcal{X}(n+1), N) \subsetneq \Gamma_1(\mathcal{X}(n), N)$ with positive probability, since the new point in $\mathcal{X}(n+1)$ has positive probability to fall closer to the subset of $T(\mathcal{Y}_3)$ that defines $\mathscr{R}_S(N)$ (e.g., $\partial(T(\mathcal{Y}_3))$). The general results in Theorems 4.8 and 4.10 and Proposition 4.9 hold for the proximity maps $\{N_S, N_{AS}, N_{PE}^r, N_{CS}^\tau\}$ also. In particular we have the following corollaries to these results.

***Corollary to Theorem 4.8:*** *Given a sequence of random variables* $X_1, X_2, \ldots \overset{iid}{\sim} \mathcal{U}(T(\mathcal{Y}_3))$*, let* $\mathcal{X}(n) := \mathcal{X}(n-1) \cup \{X_n\}$ *for* $n = 0, 1, 2, \ldots$ *with* $\mathcal{X}(0) := \emptyset$*. Then for each* $N \in \{N_S, N_{AS}, N_{PE}^r, N_{CS}^\tau\}$*,* $\Gamma_1(\mathcal{X}(n), N) \downarrow \mathscr{R}_S(N)$ *as* $n \to \infty$ *a.s., in the sense that* $\Gamma_1(\mathcal{X}(n+1), N) \subseteq \Gamma_1(\mathcal{X}(n), N)$ *and* $A(\Gamma_1(\mathcal{X}(n), N) \setminus \mathscr{R}_S(N)) \downarrow 0$ *a.s.*

***Corollary to Proposition 4.9:*** *For positive integers* $m > n$*, let* $\mathcal{X}_m$ *and* $\mathcal{X}_n$ *be two random samples from* $\mathcal{U}(T(\mathcal{Y}_3))$*. Then* $A(\Gamma_1(\mathcal{X}_m, N, M)) \leq^{ST} A(\Gamma_1(\mathcal{X}_n, N, M))$ *for proximity maps* $N \in \{N_{PE}^r, N_{CS}^\tau\}$*, for* $r \in [1, \infty)$ *and* $\tau \in (0, 1]$*.*

For $N \in \{N_S, N_{AS}\}$, the stochastic ordering of $A(\Gamma_1(\mathcal{X}_n, N, M))$ is still an open problem, although we conjecture that $A(\Gamma_1(\mathcal{X}_m, N, M)) \leq^{ST} A(\Gamma_1(\mathcal{X}_n, N, M))$ for $\mathcal{X}_m$ and $\mathcal{X}_n$ two random samples from $\mathcal{U}(T(\mathcal{Y}_3))$ with $m > n$.

***Corollary to Theorem 4.10:*** *Let* $\{\mathcal{X}_n\}_{n=1}^\infty$ *be a sequence of data sets which are iid* $\mathcal{U}(T(\mathcal{Y}_3))$*. Then* $\Gamma_1(\mathcal{X}_n, N) \xrightarrow{a.s.} \mathscr{R}_S(N)$ *as* $n \to \infty$ *for* $N \in \{N_S, N_{AS}, N_{PE}^r, N_{CS}^\tau\}$*.*

## 7.3 Expected Measure of $\Gamma_1$-Regions

Let $\lambda(\cdot)$ be the $\mathbb{R}^d$-Lebesgue measure on $\mathbb{R}^d$ with $d \geq 1$. In $\mathbb{R}$, $\lambda(\cdot)$ is the length $|\cdot|$, in $\mathbb{R}^2$, $\lambda(\cdot)$ is the area $A(\cdot)$, and in $\mathbb{R}^d$, $\lambda(\cdot)$ is the $d$-dimensional volume $V(\cdot)$. In $\mathbb{R}$ with $\mathcal{Y}_2 = \{0, 1\}$, let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(0, 1)$, and $N_S(x) = B(x, r(x))$ where $r(x) = \min(x, 1-x)$. Then,

$$\Gamma_1(\mathcal{X}_n, N_S) = [X_{n:n}/2, (1 + X_{1:n})/2] \Longrightarrow \lambda(\Gamma_1(\mathcal{X}_n, N_S)) = |\Gamma_1(\mathcal{X}_n, N_S)| = (1 + X_{1:n} - X_{n:n})/2.$$

Hence the expected length of the $\Gamma_1$-region is

$$\begin{aligned} \mathbf{E}[\lambda(\Gamma_1(\mathcal{X}_n, N_S)] &= \mathbf{E}\left[\frac{1 + X_{1:n} - X_{n:n}}{2}\right] = \frac{1 + \mathbf{E}[X_{1:n}] - \mathbf{E}[X_{n:n}]}{2} \\ &= \frac{1 + \frac{1}{n+1} - \frac{n}{n+1}}{2} = \frac{1}{n+1} \to 0 \text{ as } n \to \infty. \end{aligned}$$

In $\mathbb{R}^2$, with three non-collinear points $\mathcal{Y}_3 = \{\mathsf{y}_1, \mathsf{y}_2, \mathsf{y}_3\}$, let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(T(\mathcal{Y}_3))$. Then $A(\Gamma_1(\mathcal{X}_n, N_{PE}^r, M)) > 0$ a.s. for all $n < \infty$, $r > 3/2$, $M \in \mathbb{R}^2 \setminus \mathcal{Y}_3$. Furthermore, $A(\Gamma_1(\mathcal{X}_n, N_{CS}^{\tau=1}, M)) > 0$ a.s. for all $n < \infty$ and $M \in (T(\mathcal{Y}_3))^o$; and for $N \in \{N_S, N_{AS}\}$ $A(\Gamma_1(\mathcal{X}_n, N, M)) > 0$ a.s. for all $n < \infty$ and $M \in \mathbb{R}^2 \setminus \mathcal{Y}_3$.

The $\Gamma_1$-region, $\Gamma_1(\mathcal{X}_n, N)$, is closely related to the distribution of the domination number of the PCD associated with $N$. Hence we study the asymptotic behavior of the expected area $\mathbf{E}[A(\Gamma_1(\mathcal{X}_n, N, M)]$, as $n \to \infty$, for $N \in \{N_{AS}, N_{PE}^r, N_{CS}^\tau\}$.

In $\mathbb{R}$, $N_S$ and $N_{AS}$ are equivalent functions, with the extension that $N_{AS}$ is defined as $N_S$ for $\mathcal{X}$ points outside $(\mathsf{y}_{1:m}, \mathsf{y}_{m:m})$. In higher dimensions, determining the areas of $\Gamma_1(\mathcal{X}_n, N, M)$ for for $N \in \{N_S, N_{AS}\}$ and for general $\mathcal{X}_n$ and hence finding its expected areas are both open problems.

### 7.3.1 The Limit of Expected Area of $\Gamma_1(\mathcal{X}_n, N, M)$ for $N_{PE}^r$ and $N_{CS}^\tau$

Recall that for $N \in \{N_{PE}^r, N_{CS}^\tau\}$, $\Gamma_1(\mathcal{X}_n, N, M)$ is determined by the (closest) edge extrema $X_e(n) \in \operatorname{argmin}_{X \in \mathcal{X}_n} d(X, e)$, for $e \in \{e_1, e_2, e_3\}$. So, to find the expected area of $\Gamma_1(\mathcal{X}_n, N, M)$, we need to find the expected locus of $X_e(n)$; i.e., the expected distance of $d(X_e(n)$ from $e$. For example, for $\mathcal{X}_n$ a random

sample from a continuous distribution $F$, $\operatorname{argmin}_{X\in\mathcal{X}_n} d(X,e)$ is unique a.s., and if $d(X_e(n),e)=u$, then $X_e(n)$ falls on a line parallel to $e$ whose distance from $e$ is $u$ a.s.

**Lemma 7.4.** *Let $D_i(n) := d(X_{e_i}(n), e_i)$ for $i \in \{1,2,3\}$ and $\mathcal{X}_n$ be a random sample from $\mathcal{U}(T(\mathcal{Y}_3))$. Then $\mathbf{E}[D_i(n)] \to 0$ (i.e., the expected locus of $X_{e_i}(n)$ is on $e_i$) for each $i \in \{1,2,3\}$, as $n \to \infty$.*

**Proof:** Given $Z_i = (X_i, Y_i) \stackrel{iid}{\sim} \mathcal{U}(T(\mathcal{Y}_3))$. Then for $e = e_3$, $D_3(n) = Y_{1:n}$ (the minimum $y$-coordinate of $Z_i \in \mathcal{X}_n$). First observe that $P(Y_i \le y) = \dfrac{y\,(2\,c_2 - y)}{c_2^2}$, hence

$$F_Y(y) = \frac{y\,(2\,c_2 - y)}{c_2^2}\mathbf{I}(0 \le y < c_2) + \mathbf{I}(y \ge c_2).$$

So the pdf of $Y_i$ is $f_Y(y) = 2\,\dfrac{c_2 - y}{c_2^2}\mathbf{I}(0 \le y \le c_2)$. Then the pdf of $Y_{1:n}$ is

$$f_{1:n}(y) = 2\,n(c_2 - y)\left(1 - \frac{y\,(2\,c_2 - y)}{c_2^2}\right)^{n-1} {c_2}^{-2}\mathbf{I}(0 \le y \le c_2).$$

Therefore,

$$\mathbf{E}\,[Y_{1:n}] = \int_0^{c_2} 2\,y\,n(c_2 - y)\left(1 - \frac{y\,(2\,c_2 - y)}{c_2^2}\right)^{n-1} {c_2}^{-2} dy = \frac{c_2}{2\,n+1} \to 0, \text{ as } n \to \infty.$$

Hence $\mathbf{E}[Y_{1:n}] = \mathbf{E}[D_3(n)] \to 0$. Similarly, $\mathbf{E}[D_i(n)] \to 0$ for $i \in \{1,2\}$, as $n \to \infty$. ■

**Theorem 7.5.** *Let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(T(\mathcal{Y}_3))$ and $M \in T(\mathcal{Y}_3)^o$. For $N \in \{N_{PE}^r, N_{CS}^\tau\}$, $\mathbf{E}[A(\Gamma_1(\mathcal{X}_n, N, M))] \to A(\mathscr{R}_S(N,M))$ as $n \to \infty$.*

**Proof:** Recall that for $N \in \{N_{PE}^r, N_{CS}^\tau\}$, $\Gamma_1(\mathcal{X}_n, N, M) = \bigcap_{i=1}^3 \Gamma_1(X_{e_i}(n), N, M)$. Moreover, $\Gamma_1(\mathcal{X}_n, N, M) = \mathscr{R}_S(N,M)$ iff $X_{e_i}(n) \in e_i$ for $i \in \{1,2,3\}$. In Lemma 7.4, we have shown that expected locus of $X_e(n)$ converges to edge $e$ as $n \to \infty$. Hence the expected locus of $\partial(\Gamma_1(\mathcal{X}_n, N, M)) \cap R_M(e_i)$ converges to the $\partial(\mathscr{R}_S(N,M)) \cap R_M(e_i)$ for each $i \in \{1,2,3\}$. Hence

$$\mathbf{E}[A(\Gamma_1(\mathcal{X}_n, N, M)] \to A(\mathscr{R}_S(N,M)) \text{ as } n \to \infty. \blacksquare$$

*Remark* 7.6. In particular,

i- $\mathbf{E}\left[A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^2, M_C\right)\right)\right] \to 1/4$ as $n \to \infty$, since $\mathscr{R}_S\left(N_{PE}^2, M_C\right) = T(M_1, M_2, M_3)$.

ii- $\mathbf{E}\left[A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M\right)\right)\right] \to 0$ as $n \to \infty$ if $M \in \mathscr{T}^r$, since $\mathscr{R}_S\left(N_{PE}^r, M\right) = \emptyset$ for $M \in \mathscr{T}^r$.

iii- Furthermore, $\mathbf{E}\left[A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^{3/2}, M_C\right)\right)\right] \to 0$ since $\mathscr{R}_S\left(N_{PE}^{3/2}, M_C\right) = \{M_C\}$.

iv- For any $M \in T(\mathcal{Y}_3)^o$, $\mathbf{E}\left[A\left(\Gamma_1\left(\mathcal{X}_n, N_{CS}^\tau, M\right)\right)\right] \to 0$ as $n \to \infty$, since $\mathscr{R}_S\left(N_{CS}^\tau, M\right) = \{M\}$.

v- We also have $\mathbf{E}\left[A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M_C\right)\right)\right] \to 0$ for $r \in [1, 3/2]$ as $n \to \infty$.

vi- Furthermore, by careful geometric calculations, we get $\mathbf{E}\left[A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M_C\right)\right)\right] \to \sqrt{3}\,[1 - 3/(2\,r)]^2$ for $r \in (3/2, 2]$.

vii- $\mathbf{E}\left[A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M_C\right)\right)\right] \to \sqrt{3}/\left[4\,(1 - 3/r^2)\right]$ for $r \in (2, \infty]$, as $n \to \infty$.

We also derive the rate of convergence of $\mathbf{E}[A(\Gamma_1(\mathcal{X}_n, N_{PE}^r, M_C))]$ for $r = 3/2$.

**Theorem 7.7.** *Let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(T(\mathcal{Y}_3))$. For $r = 3/2$, the expected area of the the $\Gamma_1$-region, $\mathbf{E}\left[A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^r, M_C\right)\right)\right]$, converges to zero, at rate $O\left(n^{-2}\right)$.*

**Proof:** For $r = 3/2$ and $M = M_C$, and sufficiently large $n$, $\Gamma_1\left(X_{e_i}, N_{PE}^r\right) \cap R_{CM}(\mathsf{y}_i)$ is a triangle for $i = 1, 2, 3$ w.p. 1. See Figure 17. With the realization of the edge extrema denoted as $x_{e_i} = (x_i, y_i)$ close enough to $e_i$, for $i = 1, 2, 3$,

$\Gamma_1\left(x_{e_1}, N_{PE}^{3/2}\right) \cap R_{CM}(\mathsf{y}_1)$ is the triangle with vertices

$$\left(\frac{\sqrt{3}}{3}\,y_2 + x_2 - \frac{1}{2}, -\frac{\sqrt{3}}{18}\left(-9 + 2\sqrt{3}\,y_2 + 6\,x_2\right)\right), \left(\frac{1}{2}, \frac{\sqrt{3}}{9}\left(-\frac{9}{2} + 2\sqrt{3}\,y_2 + 6\,x_2\right)\right), \left(\frac{1}{2}, \frac{\sqrt{3}}{6}\right),$$

$\Gamma_1\left(x_{e_2}, N_{PE}^{3/2}\right) \cap R_{CM}(\mathsf{y}_2)$ is the triangle with vertices

$$\left(\frac{1}{2}, \frac{\sqrt{3}}{18}\left(3 + 4\sqrt{3}\,y_3 - 12\,x_3\right)\right), \left(\frac{1}{2} - \frac{\sqrt{3}}{3}\,y_3 + x_3, -\frac{\sqrt{3}}{18}\left(-3 + 2\sqrt{3}\,y_3 - 6\,x_3\right)\right), \left(\frac{1}{2}, \frac{\sqrt{3}}{6}\right),$$

and $\Gamma_1\left(x_{e_3}, N_{PE}^{3/2}\right) \cap R_{CM}(\mathsf{y}_3)$ is the triangle with vertices

$$\left(-\frac{\sqrt{3}}{6}\left(-\sqrt{3} + 4\,y_1\right), \frac{\sqrt{3}}{6} + \frac{2}{3}\,y_1\right), \left(\frac{1}{2}, \frac{\sqrt{3}}{6}\right), \left(\frac{\sqrt{3}}{6}\left(\sqrt{3} + 4\,y_1\right), \frac{\sqrt{3}}{6} + \frac{2}{3}\,y_1\right).$$

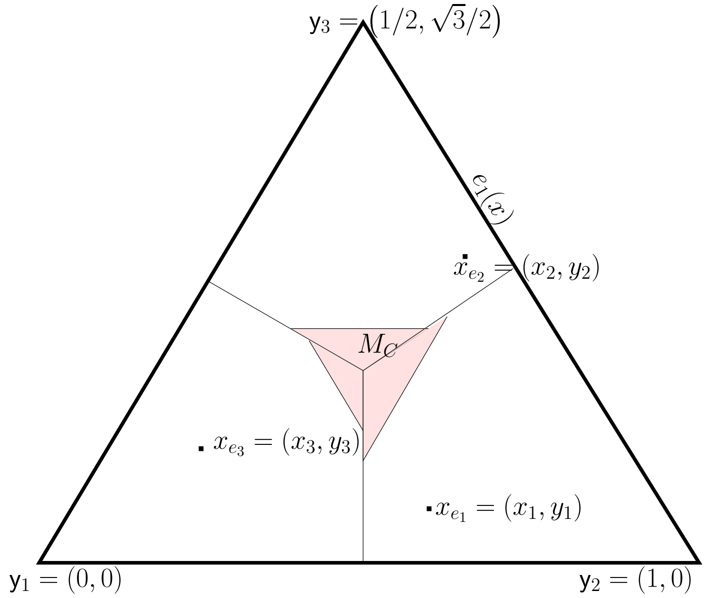


Figure 17: The shaded regions are the triangular $\Gamma_1^{3/2}(X_{e_i}) \cap R_{CM}(\mathsf{y}_i)$ regions for $i = 1, 2, 3$.

Then for sufficiently large $n$,

$$A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^{3/2}, M\right)\right) = \frac{\sqrt{3}}{27}\left(3\,x_2 - 3 + \sqrt{3}\,y_2\right)^2 + \frac{\sqrt{3}}{27}\left(-3\,x_3 + \sqrt{3}\,y_3\right)^2 + \frac{4\sqrt{3}}{9}y_1^2$$
$$= \frac{\sqrt{3}}{9}\left(3\,x_2^2 - 6\,x_2 + 2\sqrt{3}\,y_2\,x_2 - 2\sqrt{3}\,y_2 + y_2^2 + 3 + y_3^2 - 2\sqrt{3}\,y_3\,x_3 + 3\,x_3^2 + 4\,y_1^2\right).$$

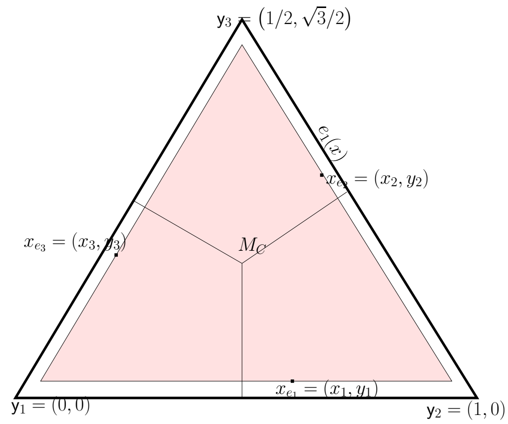


Figure 18: The figure for the joint pdf of $X_{e_i}$. The shaded region is $T(\zeta)$.

To find the expected area, we need the joint density of the $X_{e_i}$. The edge extrema are all distinct with probability 1 as $n \to \infty$ (see Theorem 6.4). Let $T(\zeta)$ be the triangle formed by the lines at $x_{e_i}$ parallel to $e_i$ for $i = 1, 2, 3$ where $\zeta = (x_1, y_1, x_2, y_2, x_3, y_3)$. See Figure 18.

Then the asymptotically accurate joint pdf of $X_{e_i}$ (Ceyhan and Priebe (2007)) is

$$\begin{aligned}
f_3(\zeta) &= n(n-1)(n-2)\left(\frac{A(T(\zeta))}{A(T(\mathcal{Y}_3))}\right)^{n-3} \frac{1}{A(T(\mathcal{Y}_3))^3} \\
&= n(n-1)(n-2)\left(\sqrt{3}/36\left(-2\sqrt{3}\,y_1 + \sqrt{3}\,y_3 - 3\,x_3 + \sqrt{3}\,y_2 + 3\,x_2\right)^2\right)^{n-3} \Big/ \left(\sqrt{3}/4\right)^n .
\end{aligned}$$

with the support $D_S = \left\{\zeta \in \Re^6 : (x_i, y_i)\text{'s are distinct}\right\}$.

Then for sufficiently large $n$,

$$\begin{aligned}
\mathbf{E}\left[A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^{3/2}, M\right)\right)\right] &\sim \int A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^{3/2}\right)\right) f_3(\zeta)\, d\zeta \\
&= \int A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^{3/2}\right)\right) n(n-1)(n-2)\left(\frac{A(T(\zeta))}{A(T(\mathcal{Y}_3))}\right)^{n-3} \frac{1}{A(T(\mathcal{Y}_3))^3}\, d\zeta.
\end{aligned}$$

Let $G(\zeta) = A(T(\zeta))/A(T(\mathcal{Y}_3))$. Notice that the integrand is critical when $x_{e_i} \in e_i$ for $i = 1, 2, 3$, since $G(\zeta)$ when $x_{e_i} \in e_i$ for each $i = 1, 2, 3$. So we make the change of variables $y_1 = z_1$, $y_2 = \sqrt{3}\,(1 - x_2) - z_2$, and $y_3 = \sqrt{3}\,x_3 - z_3$, then $G(\zeta)$ and $A(\Gamma_1(\mathcal{X}_n, N))$ becomes

$$G(\zeta) = G(z_1, z_2, z_3) = \left(2\,z_1 + z_3 - \sqrt{3} + z_2\right)^2 \Big/ 3 \text{ and } A(\Gamma_1(\mathcal{X}_n, N)) = \sqrt{3}\,\left(z_2^2 + z_3^2 + 4\,z_1^2\right)/9,$$

respectively. Hence the integrand does not depend on $x_1, x_2, x_3$ and integrating with respect to $x_1, x_2, x_3$ yields a constant $K$. Now, the integrand is critical at $(z_1, z_2, z_3) = (0, 0, 0)$, since $G(0, 0, 0) = 1$. So let $E_u^\varepsilon$ be the event that $0 \le z_i \le \varepsilon$ for $i = 1, 2, 3$ for $\varepsilon > 0$ small enough. Then making the change of variables $z_i = w_i/n$ for $i = 1, 2, 3$, we get $A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^{3/2}\right)\right) = O\left(n^{-2}\right)$ and $G(z_1, z_2, z_3)$ becomes

$G(w_1,w_2,w_3) = 1 - \frac{1}{n}\left(2/\sqrt{3}\,(2\,w_1+w_2+w_3)\right) + O\left(n^{-2}\right)$, hence

$$\begin{aligned}
&\mathbf{E}\left[A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^{3/2}, M\right)\right)\right] \sim \\
&K\int_0^{n\varepsilon}\int_0^{n\varepsilon}\int_0^{n\varepsilon} A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^{3/2}\right)\right) n(n-1)(n-2)\frac{1}{n^3}\,G(w_1,w_2,w_3)^{n-3} dw_1 dw_2 dw_3, \\
&\text{letting } n\to\infty \\
&\approx K\int_0^{\infty}\int_0^{\infty}\int_0^{\infty} O\left(n^{-2}\right)\exp\left(-2/\sqrt{3}\,(2\,w_1+w_2+w_3)\right) dw_1 dw_2 dw_3 = O\left(n^{-2}\right),
\end{aligned}$$

since $\int_0^\infty\int_0^\infty\int_0^\infty \exp\left(-2/\sqrt{3}\,(2\,w_1+w_2+w_3)\right) dw_1dw_2dw_3 = 3\sqrt{3}/16$ which is a finite constant. Hence $\mathbf{E}\left[A\left(\Gamma_1\left(\mathcal{X}_n, N_{PE}^{3/2}, M\right)\right)\right] \to 0$ as $n\to\infty$ at the rate $O\left(n^{-2}\right)$. $\blacksquare$

# 8 $\Gamma_k$-Regions for Proximity Maps in $T\left(\mathcal{Y}_3\right)$

We can also define the regions associated with $\gamma_n(N) = k$ for $k \le n$.

In $\mathbb{R}$ with $\mathcal{Y}_2 = \{0,1\}$, $\gamma_n(N) \le 2$, hence we can only define $\Gamma_2$-regions. Recall that $\gamma_n(N) = 2$ iff $\mathcal{X}_n \cap \left[\frac{x_{n:n}}{2}, \frac{1+x_{1:n}}{2}\right] = \emptyset$ iff $\mathcal{X}_n \subset [0,1]\setminus\Gamma_1\left(\mathcal{X}_n,N\right)$. So

$$\begin{aligned}
\Gamma_2\left(\mathcal{X}_n,N\right) &= \left\{(x,y)\in[0,1]^2:\ \mathcal{X}_n\subset N(x)\cup N(y);\ x,y\notin\Gamma_1\left(\mathcal{X}_n,N\right)\right\}\\
&= \left\{(x,y)\in[0,1]^2\setminus\Gamma_1\left(\mathcal{X}_n,N\right)^2:\ \mathcal{X}_n\subset N(x)\cup N(y)\right\}.
\end{aligned}$$

Notice that $\Gamma_2\left(\mathcal{X}_n,N\right)\subseteq[0,1]^2$. Let

$$\widetilde{x}_1 := \operatorname{argmin}_{x\in\mathcal{X}_n\cap(0,1/2)}(1/2-x) \text{ and } \widetilde{x}_2 := \operatorname{argmin}_{x\in\mathcal{X}_n\cap(1/2,1)}(x-1/2),$$

then $\gamma_n(N)=2$ iff $\widetilde{x}_1,\widetilde{x}_2\notin\Gamma_1\left(\mathcal{X}_n,N\right)$. In such a case $\mathcal{X}_n\subset N(\widetilde{x}_1)\cup N(\widetilde{x}_2)$ by construction.

In general,

**Definition 8.1.** The $\Gamma_2$*-region* for proximity map $N_{\mathcal{Y}}(\cdot)$ and set $B\subset\Omega$ is $\Gamma_2(B,N)=\{(x,y)\in[\Omega\setminus\Gamma_1(B)]^2: B\subseteq N_{\mathcal{Y}}(x)\cup N_{\mathcal{Y}}(y)\}$. In general, $\Gamma_k$*-region* for proximity map $N_{\mathcal{Y}}(\cdot)$ and set $B\subset\Omega$ for $k=1,2,\ldots,n$ is

$$\begin{aligned}
\Gamma_k(B,N) = \Bigg\{&(x_1,x_2,\ldots,x_k)\in\Omega^k : B\subseteq\bigcup_{i=1}^k N_{\mathcal{Y}}(x_i) \text{ and all possible } m\text{-permutations } (u_1,u_2,\ldots,u_m)\\
&\text{of } (x_1,x_2,\ldots,x_k) \text{ satisfy } (u_1,u_2,\ldots,u_m)\notin\Gamma_m(B,N) \text{ for each } m=1,2,\ldots,k-1\Bigg\}.\square
\end{aligned}$$

Note that $\Gamma_k$-regions are defined for $k\le n$ and they might be empty. Moreover, $\Gamma_k$-regions are in $\Omega^k$, not in $\Omega$.

Let $\left\{ n \atop m \right\}$ denote the Stirling partition number for a set of size $n$ into $m$ blocks and let $\left\{ A \atop m \right\}$ denote all Stirling partitions of a set $A$ into $m$ blocks; that is,

$$\left\{ A \atop m \right\} := \left\{\{\mathcal{B}_1,\mathcal{B}_2,\ldots,\mathcal{B}_m\}:\ \mathcal{B}_i\ne\emptyset, \mathcal{B}_i\subset A;\ A=\bigcup_i\mathcal{B}_i\right\}.$$

In particular, $\left\{ \mathcal{X}_n \atop 2 \right\}$ is the unordered pair of blocks $\mathcal{B}_1$ and $\mathcal{B}_2$ such that $\mathcal{B}_i\ne\emptyset$ and $\mathcal{B}_i\subset\mathcal{X}_n$ for $i=1,2$, and $\mathcal{B}_1\cup\mathcal{B}_2=\mathcal{X}_n$. Note that $\mathcal{B}_2=\mathcal{X}_n\setminus\mathcal{B}_1$. Then

**Proposition 8.2.** $\Gamma_2(\mathcal{X}_n, N) = \bigcup_{\{\mathcal{B}_1,\mathcal{B}_2\}\in\left\{ {\mathcal{X}_n \atop 2} \right\}} [\Gamma_1(\mathcal{B}_1, N) \times \Gamma_1(\mathcal{B}_2, N)] \setminus \Gamma_1(\mathcal{X}_n, N)^2$ *for any nonempty Stirling blocks* $\mathcal{B}_1$ *and* $\mathcal{B}_2$ *in* $\left\{ {\mathcal{X}_n \atop 2} \right\}$.

**Proof:** Given $\mathcal{X}_n$, suppose $(u, v) \in \Gamma_2(\mathcal{X}_n, N)$, then $\mathcal{X}_n \subset N(u) \cup N(v)$ and $u, v \notin \Gamma_1(\mathcal{X}_n, N)$. Let $\mathcal{B}_1 = \mathcal{X}_n \cap N(u)$, and $\mathcal{B}_2 = [\mathcal{X}_n \cap N(v)] \setminus N(u)$. Then $\mathcal{B}_1$ and $\mathcal{B}_2$ are two Stirling blocks in $\left\{ {\mathcal{X}_n \atop 2} \right\}$. Hence $\mathcal{B}_1 \subset N(u) \Rightarrow u \in \Gamma_1(\mathcal{B}_1, N) \setminus \Gamma_1(\mathcal{X}_n, N)$ and $\mathcal{B}_2 \subset N(v) \Rightarrow v \in \Gamma_1(\mathcal{B}_2, N) \setminus \Gamma_1(\mathcal{X}_n, N)$, hence $\Gamma_2(\mathcal{X}_n, N) \subseteq \bigcup_{\{\mathcal{B}_1,\mathcal{B}_2\}\in\left\{ {\mathcal{X}_n \atop 2} \right\}} [\Gamma_1(\mathcal{B}_1, N) \times \Gamma_1(\mathcal{B}_2, N)] \setminus (\Gamma_1(\mathcal{X}_n, N))^2$. The other direction is trivial, hence the desired result follows. ■

In $\mathbb{R}$ with $N = N_S$, we can exploit the natural ordering available.

**Proposition 8.3.** *In* $\mathbb{R}$ *with* $\mathcal{Y}_2 = \{0, 1\}$, $\Gamma_2(\mathcal{X}_n, N_S) = \displaystyle\bigcup_{k=1}^{n-1} \left( \frac{x_{k:n}}{2}, \frac{x_{n:n}}{2} \right) \times \left( \frac{1 + x_{1:n}}{2}, \frac{1 + x_{(k+1):n}}{2} \right)$.

**Proof:** Recall that $\Gamma_2(\mathcal{X}_n, N_S) = \bigcup_{\{\mathcal{B}_1,\mathcal{B}_2\}\in\left\{ {\mathcal{X}_n \atop 2} \right\}} [\Gamma_1(\mathcal{B}_1, N_S) \times \Gamma_1(\mathcal{B}_2, N_S)] \setminus \Gamma_1(\mathcal{X}_n, N_S)^2$. Let $\{\mathcal{B}_1, \mathcal{B}_2\}$ be a Stirling partition in $\left\{ {\mathcal{X}_n \atop 2} \right\}$. First observe that $\mathcal{B}_1 = \{x_{1:n}, \ldots, x_{k:n}\}$ and $\mathcal{B}_2 = \{x_{(k+1):n}, \ldots, x_{n:n}\}$ forms a Stirling partition of $\mathcal{X}_n$, and

$$\Gamma_1(\{x_{1:n}, \ldots, x_{k:n}\}, N_S) \setminus \Gamma_1(\mathcal{X}_n, N_S) = \left[ \frac{x_{k:n}}{2}, \frac{1 + x_{1:n}}{2} \right] \setminus \left[ \frac{x_{n:n}}{2}, \frac{1 + x_{1:n}}{2} \right] = \left[ \frac{x_{k:n}}{2}, \frac{x_{n:n}}{2} \right)$$

and

$$\Gamma_1\left(\{x_{(k+1)}, \ldots, x_{n:n}\}, N_S\right) \setminus \Gamma_1(\mathcal{X}_n, N_S) = \left[ \frac{x_{n:n}}{2}, \frac{1 + x_{(k+1)}}{2} \right] \setminus \left[ \frac{x_{n:n}}{2}, \frac{1 + x_{1:n}}{2} \right] = \left( \frac{1 + x_{1:n}}{2}, \frac{1 + x_{(k+1)}}{2} \right].$$

Furthermore,

$$\Gamma_1(\{x_{1:n}, \ldots, x_{k:n}\}, N_S) \times \Gamma_1\left(\{x_{(k+1)}, \ldots, x_{n:n}\}, N_S\right) \supseteq \bigcup_{\{\mathcal{B}_1,\mathcal{B}_2\}\in\Lambda} \Gamma_1(\mathcal{B}_1, N_S) \times \Gamma_1(\mathcal{B}_2, N_S) \tag{4}$$

where

$$\Lambda = \left\{ \{\mathcal{B}_1, \mathcal{B}_2\} \in \left\{ {\mathcal{X}_n \atop 2} \right\} : \; x_{1:n} \in \mathcal{B}_1, \; x_{n:n} \in \mathcal{B}_2, \; \max(\mathcal{B}_1) = x_{k:n}, \; \min(\mathcal{B}_2) = x_{(l)}, \; l < k \right\},$$

since the left hand side in Equation (4) is $\left[ \frac{x_{k:n}}{2}, \frac{x_{n:n}}{2} \right] \times \left[ \frac{1 + x_{1:n}}{2}, \frac{1 + x_{(k-1)}}{2} \right]$. Hence the desired result follows. ■

Similarly, for $N^r_{PE}$,

$$\Gamma_2(\mathcal{X}_n, N^r_{PE}) = \bigcup_{\{\mathcal{B}_1,\mathcal{B}_2\}\in\left\{ {\mathcal{X}_n \atop 2} \right\}} [\Gamma_1(\mathcal{B}_1, N^r_{PE}, M) \times \Gamma_1(\mathcal{B}_2, N^r_{PE}, M)] \setminus \Gamma_1(\mathcal{X}_n, N^r_{PE}, M)^2.$$

Note that $\Gamma_1(\mathcal{B}_i, N^r_{PE}, M)$ is determined by the edge extrema in $\mathcal{B}_i$, $i = 1, 2$. Furthermore, if $(u, v) \in \Gamma_2(\mathcal{X}_n, N^r_{PE}, M)$, then $(u, v) \notin R_M(\mathsf{y})^2$, since either $N^r_{PE}(u) \subseteq N^r_{PE}(v)$ or $N^r_{PE}(v) \subseteq N^r_{PE}(u)$ should hold if $(u, v) \in R_M(\mathsf{y})^2$.

For $N^\tau_{CS}$,

$$\Gamma_2(\mathcal{X}_n, N^\tau_{CS}, M) = \bigcup_{\{\mathcal{B}_1,\mathcal{B}_2\}\in\left\{ {\mathcal{X}_n \atop 2} \right\}} [\Gamma_1(\mathcal{B}_1, N^\tau_{CS}, M) \times \Gamma_1(\mathcal{B}_2, N^\tau_{CS}, M)] \setminus \Gamma_1(\mathcal{X}_n, N^\tau_{CS}, M)^2.$$

Note that $\Gamma_1(\mathcal{B}_i, N_{CS}^{\tau}, M)$ is determined by the edge extrema of $\mathcal{B}_i$, $i = 1, 2$. But if $(u, v) \in \Gamma_2(\mathcal{X}_n, N_{CS}^{\tau}, M)$, then $(u, v) \in R_M(e)^2$ can hold. In these proximity regions, not all edge extrema should fall in the same partition set for $\Gamma_k(\mathcal{X}_n, N_{CS}^{\tau}, M)$ to be nonempty.

For any proximity map $N$,

$$\begin{aligned} &P(\gamma_n(N) = 2) = P(\mathcal{X}_n^2 \cap \Gamma_2(\mathcal{X}_n, N) \neq \emptyset,\ \gamma_n(N) \neq 1) \\ &= P\left(\mathcal{X}_n^2 \bigcap \left[\bigcup_{\{\mathcal{B}_1,\mathcal{B}_2\} \in \left\{\substack{\mathcal{X}_n \\ 2}\right\}} [\Gamma_1(\mathcal{B}_1, N) \times \Gamma_1(\mathcal{B}_2, N)] \setminus (\Gamma_1(\mathcal{X}_n, N))^2\right] \neq \emptyset,\ \mathcal{X}_n \cap \Gamma_1(\mathcal{X}_n, N) = \emptyset\right). \end{aligned}$$

A more compact way to write this is as

$$P(\gamma_n(N) > 2) = P\left(\mathcal{X}_n^2 \cap \Gamma_{\leq 2}(\mathcal{X}_n, N) = \emptyset\right)$$

where $\Gamma_{\leq 2}(\mathcal{X}_n, N) := \bigcup_{\{\mathcal{B}_1,\mathcal{B}_2\} \in \left\{\substack{\mathcal{X}_n \\ 2}\right\}} \Gamma_1(\mathcal{B}_1, N) \times \Gamma_1(\mathcal{B}_2, N)$.

Furthermore, for $k \geq 3$, the $\Gamma_k$-regions are defined similarly as

$$\Gamma_k(\mathcal{X}_n, N) = \bigcup_{\{\mathcal{B}_1,\ldots,\mathcal{B}_k\} \in \left\{\substack{\mathcal{X}_n \\ k}\right\}} [\Gamma_1(\mathcal{B}_1, N) \times \ldots \times \Gamma_1(\mathcal{B}_k, N)] \setminus \Gamma_1(\mathcal{X}_n, N)^k.$$

Hence,

$$\begin{aligned} P(\gamma_n(N) = k) &= P\left(\mathcal{X}_n^k \cap \Gamma_k(\mathcal{X}_n, N) \neq \emptyset, \gamma_n(N) > k-1\right) \\ &= P\left(\mathcal{X}_n^k \cap \left[\bigcup_{\{\mathcal{B}_1,\ldots,\mathcal{B}_k\} \in \left\{\substack{\mathcal{X}_n \\ k}\right\}} \Gamma_1(\mathcal{B}_1, N) \ldots \times \Gamma_1(\mathcal{B}_k, N) \setminus \Gamma_1(\mathcal{X}_n, N)^k\right] \neq \emptyset\right). \end{aligned}$$

A more compact way to write this is as

$$P(\gamma_n(N) > k) = P\left(\mathcal{X}_n^k \cap \Gamma_{\leq k}(\mathcal{X}_n, N) \neq \emptyset\right)$$

where $\Gamma_{\leq k}(\mathcal{X}_n, N) := \bigcup_{\{\mathcal{B}_1,\ldots,\mathcal{B}_k\} \in \left\{\substack{\mathcal{X}_n \\ k}\right\}} \Gamma_1(\mathcal{B}_1, N) \times \ldots \times \Gamma_1(\mathcal{B}_k, N)$.

# 9 $\kappa$-Values for the Proximity Maps in $T(\mathcal{Y}_3)$

Recall that the domination number, $\gamma_n(N)$ is the cardinality of a minimum dominating set of the PCD based on $N$. So by definition, $\gamma_n(N) \leq n$. We will seek an a.s. least upper bound for $\gamma_n(N)$ which suggests the following concept.

**Definition 9.1.** Let $\mathcal{X}_n$ be a random sample from $F$ on $T(\mathcal{Y}_3)$ and let $\gamma_n(N)$ be the domination number for the PCD based on a proximity map $N$. The general a.s. least upper bound for $\gamma_n(N)$ that works for all $n \geq 1$ is called the *$\kappa$-value*; i.e., $\kappa_n(N) := \min\{k(n) : \gamma(\mathcal{X}_n, N) \leq k(n) \text{ a.s. for all } n \geq 1\}$. $\square$

It is more desirable to have a $\kappa$-value that is independent of $n$. Further, if $\kappa_n(N) = \kappa$ exists for $N$ and is independent of $n$, then the domination number has the following discrete probability mass function:

$$\gamma(\mathcal{X}_n, N) = \begin{cases} 1 & \text{w.p.} \quad p_1, \\ 2 & \text{w.p.} \quad p_2, \\ \vdots & \quad \vdots \qquad \vdots, \\ \kappa & \text{w.p.} \quad p_\kappa. \end{cases}$$

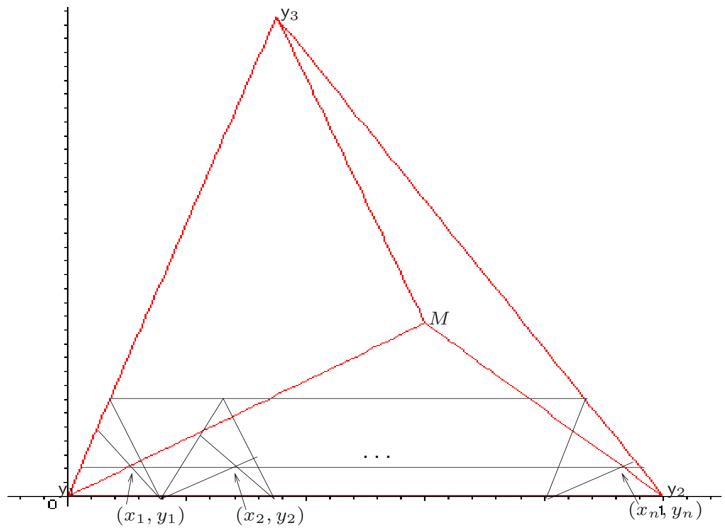


Figure 19: The figure for $\kappa_n\left(N_{CS}^{\tau}\right) = n$.

In $\mathbb{R}$ with $\mathcal{Y}_2 = \{0,1\}$, for $\mathcal{X}_n$ a random sample from $\mathcal{U}(0,1)$, $\gamma_n(N) \leq 2$ with equality holding with positive probability for $N \in \{N_S, N_{AS}\}$. Hence $\kappa_n(N_S) = 2$. But in $\mathbb{R}^d$ with $d > 1$, finding $\kappa_n(N)$ for $N \in \{N_S, N_{AS}\}$ is an open problem. Next, we investigate the $\kappa$-values for $\{N_{AS}, N_{PE}^r, N_{CS}^{\tau}\}$ in $\mathbb{R}^2$.

**Theorem 9.2.** *Let $\mathcal{X}_n$ be a random sample from $\mathcal{U}(T(\mathcal{Y}_3))$, and $M \in \mathbb{R}^2 \setminus \mathcal{Y}_3$. For $N_{PE}^r(\cdot, M)$, $\kappa_n\left(N_{PE}^r\right) = 3$.*

**Proof:** For $N_{PE}^r(\cdot, M)$, pick the point closest to edge $e_i$ in vertex region $R_M(\mathsf{y}_i)$; that is, pick $U_i \in \operatorname{argmin}_{X \in \mathcal{X}_n \cap R_M(\mathsf{y}_i)} d(X, e_i) = \operatorname{argmax}_{X \in \mathcal{X}_n \cap R_M(\mathsf{y}_i)} d(\ell(\mathsf{y}, X), \mathsf{y}_i)$ in the vertex region for which $\mathcal{X}_n \cap R_M(\mathsf{y}_i) \neq \emptyset$ for $i \in \{1,2,3\}$ (note that as $n \to \infty$, $\mathcal{X}_n \cap R_M(\mathsf{y}_i) \neq \emptyset$ for all $i \in \{1,2,3\}$ a.s., and also $U_i$ is unique a.s. for each $i$, since $X$ is from $\mathcal{U}(T(\mathcal{Y}_3))$). Then $\mathcal{X}_n \cap R_M(\mathsf{y}_i) \subset N_{PE}^r(U_i, M)$. Hence $\mathcal{X}_n \subset \bigcup_{i=1}^3 N_{PE}^r(U_i, M)$. So $\gamma_n\left(N_{PE}^r, M_C\right) \leq 3$ with equality holding with positive probability. Thus $\kappa_n\left(N_{PE}^r\right) = 3$. ■

There is no least upper bound for $\gamma_n(N_{CS}^{\tau}, M)$ that works for all $n > 0$ as shown below.

**Theorem 9.3.** *Let $\mathcal{X}_n$ be random sample from $\mathcal{U}(T(\mathcal{Y}_3))$. Then $\gamma_n(N_{CS}^{\tau}, M) = n$ holds with positive probability for all $\tau \in [0,1]$.*

**Proof:** For $\tau = 0$, the result follows trivially. For $\tau = 1$, we will prove the theorem by showing that there is a union of $n$ regions of positive area in $T(\mathcal{Y}_3)$, so that $u \in N_{CS}^{\tau=1}(v, M)$ iff $u = v$, for any $u, v \in \mathcal{X}_n$. Let $M = (m_1, m_2) \in T(\mathcal{Y}_3)^o$. In $R_M(e_3)$ locate $n$ triangles evenly on $e_3$ with base length $1/n$ and similar to $T(\mathcal{Y}_3)$ (with similarity ratio $1/n$). See also Figure 19. Then locate $n$ points in each triangle at $z_i = (x_i, y_i)$ such that $(x_i, y_i)$ is the same type of center of $T_i$ as $M$ is of $T(\mathcal{Y}_3)$. Then using the similarity ratio of $N_{CS}^{\tau=1}(z_i, M)$ to $T(\mathcal{Y}_3)$, namely, $y_i/m_2 = 1/n$, we get $y_i = m_2/n$ for all $i = 1, 2, \dots, n$. Moreover, $x_i - x_{i-1} = 1/n$ for $i = 2, 3, \dots, n$ with $x_1 = m_1/n$ and $x_n = 1 - (1 - m_1)/n$. Then $(x_i, y_i) \in N_{CS}^{\tau=1}((x_j, y_j), M)$ iff $i = j$. Furthermore, for sufficiently small $\varepsilon > 0$, the same holds for the $\varepsilon$ neighborhood of each $z_i = (x_i, y_i)$. That is, $N_{CS}^{\tau=1}(x, M) \cap B(z_j, \varepsilon) = \emptyset$ for all $x \in B(z_i, \varepsilon)$ for any distinct pair $i, j \in \{1, 2, \dots, n\}$, and probability of $\mathcal{X}_n$ being composed of $n$ points one from each $B(z_i, \varepsilon)$ is positive. Then $\gamma_n(N_{CS}^{\tau}, M) = n$ holds with positive probability. The result for $\tau \in (0,1)$ follows similarly. ■

## 9.1 Characterization of Proximity Maps Using $\kappa$-Values

Notice that for the proximity maps we have considered, $\kappa_n\left(N_{PE}^r\right) = 3$ and $\kappa_n\left(N_{CS}^\tau\right) = n$. One property of proximity maps that makes $\kappa_n(N) < n$ is that probability of having an $\mathcal{X}_n$ for which $N(X) \cap \mathcal{X}_n = \{X\}$ for all $X \in \mathcal{X}_n$ is zero.

Below we state a condition for $\kappa_n(N(\cdot, M)) \leq 3$ for $N(\cdot, M)$ defined with $M$-vertex regions.

**Theorem 9.4.** *Suppose $N(\cdot, M)$ is defined with $M$-vertex regions with $M \in \mathbb{R}^2 \setminus \mathcal{Y}_3$ and $N(x, M)$ gets larger as $d(\ell(\mathsf{y}, x), \mathsf{y})$ increases for $x \in R_M(\mathsf{y})$ in the sense that $N(x, M) \subseteq N(z, M)$ for all $x, z \in R_M(\mathsf{y})$ when $d(\ell(\mathsf{y}, x), \mathsf{y}) \leq d(\ell(\mathsf{y}, z), \mathsf{y})$. Then $\kappa_n(N) \leq 3$.*

**Proof:** When $\mathcal{X}_n \cap R_M(\mathsf{y}_i) \neq \emptyset$, pick one of the points $U_i(n) \in \operatorname{argmax}_{X \in \mathcal{X}_n \cap R_M(\mathsf{y}_i)} d(\ell(\mathsf{y}_i, X), \mathsf{y}_i)$, then $\mathcal{X}_n \cap R_M(\mathsf{y}_i) \subset N(U_i(n))$ for each $i \in \{1, 2, 3\}$. So $\gamma\left(\mathcal{X}_n, N, M\right) \leq 3$, and hence $\kappa_n(N) \leq 3$. ■

Notice that $N_{PE}^r$ satisfies the conditions of Theorem 9.1.

**Theorem 9.5.** *Suppose $N(\cdot, M)$ is defined with $M$-edge regions and $N(x, M)$ gets larger as $d(x, e)$ increases for $x \in R_M(e)$ in the sense that $N(x, M) \subseteq N(y, M)$ for all $x, y \in R_M(e)$ when $d(x, e) \leq d(y, e)$. Then $\kappa_n(N) \leq 3$.*

**Proof:** When $\mathcal{X}_n \cap R_M(e_i) \neq \emptyset$, pick one of the points $U_{e_i}(n) \in \operatorname{argmax}_{X \in \mathcal{X}_n \cap R_M(e_i)} d(X, e_i)$. Then $\mathcal{X}_n \cap R_M(e_i) \subset N(U_e(n))$ for each $i \in \{1, 2, 3\}$. So $\gamma\left(\mathcal{X}_n, N, M\right) \leq 3$, and hence $\kappa_n(N) = 3$. ■

In Theorems 9.1 and 9.1, we have an upper bound for $\kappa_n(N)$. To determine an exact value for $\kappa_n(N)$ we need further restrictions. Let $A_1 := \{x \in T\left(\mathcal{Y}_3\right) : \mu(N(x, M)) = \mu(T\left(\mathcal{Y}_3\right))\}$ and $A_2 := \{(x, y) \in T\left(\mathcal{Y}_3\right) \times T\left(\mathcal{Y}_3\right) : \mu(N(x, M) \cup N(y, M)) = \mu(T\left(\mathcal{Y}_3\right))\}$. If in addition to the hypothesis of Theorem (and ) we have $A_1$ and $A_2$ have zero measure (e.g., zero area for continuous distributions with support in $T\left(\mathcal{Y}_3\right)$) then $\kappa_n(N) = 3$ would hold.

# 10 Discussion and Conclusions

In this article, we provide a probabilistic characterization of proximity maps, associated regions, and digraphs and related quantities. In particular, we discuss the probabilistic behavior of proximity regions, superset regions, and $\Gamma_1$-regions; construct digraphs (proximity catch digraphs(PCDs)) and investigate related quantities such as domination number, $\eta$ and *kappa* values. We also provide auxiliary tools such as vertex and edge regions for the construction of proximity regions.

Although $\Gamma_1$-regions and superset regions were introduced before (Ceyhan and Priebe (2005),Ceyhan et al. (2006), and Ceyhan and Priebe (2007)) a thorough investigation is only performed in this article. $G_1$-regions are a sort of a "dual" of proximity regions and are associated with domination number being equal to 1. We provide a probabilistic characterization of $\Gamma_1$-regions for general proximity maps $N$ and data points from distribution $F$. We also extend this concept by introducing $\Gamma_k$-regions, which are associated with domination number being equal to $k$.

We introduce the quantities related to $\Gamma_1$-regions and domination number, namely, $\eta$-values and $\kappa$-values. $\eta$-value is the minimum number of points in a set required to determine the $\Gamma_1$-region for that set. We determine some general conditions that make $\eta_n(N) \leq 3$ for data in the triangle $T\left(\mathcal{Y}_3\right)$. $\kappa$-value is the a.s. least upper bound for the domination number of the PCDs. We also determine some general conditions that make $\kappa_n(N) \leq 3$ for data in the triangle $T\left(\mathcal{Y}_3\right)$.

We provide two PCD families, namely proportional-edge PCDs (Ceyhan et al. (2006) and central similarity PCDs (Ceyhan et al. (2007)) as illustrative examples. We discuss the construction of proximity regions and $\Gamma_1$-regions for these PCDs. Furthermore, we calculate the expected area of $\Gamma_1$-regions for these PCDs in the limit.

Determining $\Gamma_1$-regions, $\eta$ and $\kappa$ values for spherical and arc-slice PCDs contain many open problems and are subjects of ongoing research.

With the above characterizations, given another PCD, then we can determine how it behaves in terms of $\Gamma_1$-regions and related quantities; in particular we can determine a.s. least upper bounds for the domination number of the new PCD.

## Acknowledgments

This research was supported by the research agency TUBITAK via the Kariyer Project # 107T647.

## References

Billingsley, P. (1995). *Probability and Measure*. Wiley-Interscience, New York, NY.

Boland, P. J., Singh, H., and Cukic, B. (2004). The stochastic precedence ordering with applications in sampling and testing. *Journal of Applied Probability*, 41(1):73–82.

Ceyhan, E. (2004). *An Investigation of Proximity Catch Digraphs in Delaunay Tessellations*. PhD thesis, The Johns Hopkins University, Baltimore, MD, 21218.

Ceyhan, E. (2009). Extension of one-dimensional proximity regions to higher dimensions. arXiv:0902.1306v1 [math.MG]. Technical Report # KU-EC-08-2.

Ceyhan, E. and Priebe, C. E. (2005). The use of domination number of a random proximity catch digraph for testing spatial patterns of segregation and association. *Statistics & Probability Letters*, 73:37–50.

Ceyhan, E. and Priebe, C. E. (2007). On the distribution of the domination number of a new family of parametrized random digraphs. *Model Assisted Statistics and Applications*, 1(4):231–255.

Ceyhan, E., Priebe, C. E., and Marchette, D. J. (2007). A new family of random graphs for testing spatial segregation. *Canadian Journal of Statistics*, 35(1):27–50.

Ceyhan, E., Priebe, C. E., and Wierman, J. C. (2006). Relative density of the random $r$-factor proximity catch digraphs for testing spatial patterns of segregation and association. *Computational Statistics & Data Analysis*, 50(8):1925–1964.

Chartrand, G. and Lesniak, L. (1996). *Graphs & Digraphs*. Chapman & Hall/CRC Press LLC, Florida.

DeVinney, J. (2003). *The Class Cover Problem and its Applications in Pattern Recognition*. PhD thesis, The Johns Hopkins University, Baltimore, MD, 21218.

DeVinney, J. and Priebe, C. E. (2006). A new family of proximity graphs: Class cover catch digraphs. *Discrete Applied Mathematics*, 154(14):1975–1982.

DeVinney, J., Priebe, C. E., Marchette, D. J., and Socolinsky, D. (2002). Random walks and catch digraphs in classification. `http://www.galaxy.gmu.edu/interface/I02/I2002Proceedings/DeVinneyJason/DeVinneyJason.paper.pdf`. Proceedings of the $34^{\text{th}}$ Symposium on the Interface: Computing Science and Statistics, Vol. 34.

DeVinney, J. and Wierman, J. C. (2003). A SLLN for a one-dimensional class cover problem. *Statistics & Probability Letters*, 59(4):425–435.

Devroye, L., Gyorfi, L., and Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition*. Springer Verlag, New York.

Garfinkel, R. S. and Nemhauser, G. L. (1972). *Integer Programming*. John Wiley & Sons, New York.

Janson, S., Łuczak, T., and Rucinński, A. (2000). *Random Graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., New York.

Jaromczyk, J. W. and Toussaint, G. T. (1992). Relative neighborhood graphs and their relatives. *Proceedings of IEEE*, 80:1502–1517.

Karr, A. F. (1992). *Probability*. Springer-Verlag, New York.

Lee, C. (1998). Domination in digraphs. *Journal of Korean Mathematical Society*, 4:843–853.

Liu, R. Y., Parelius, J. M., and Singh, K. (1999). Multivariate analysis by data depth: Descriptive statistics graphics and inference (with discussion). *The Annals of Statistics*, 27:783–858.

Marchette, D. J. and Priebe, C. E. (2003). Characterizing the scale dimension of a high dimensional classification problem. *Pattern Recognition*, 36(1):45–60.

Okabe, A., Boots, B., and Sugihara, K. (2000). *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. Wiley.

Paterson, M. S. and Yao, F. F. (1992). On nearest neighbor graphs. In *Proceedings of* $19^{th}$ *Int. Coll. Automata, Languages and Programming, Springer LNCS*, volume 623, pages 416–426.

Priebe, C. E., DeVinney, J. G., and Marchette, D. J. (2001). On the distribution of the domination number of random class cover catch digraphs. *Statistics & Probability Letters*, 55:239–246.

Priebe, C. E., Marchette, D. J., DeVinney, J., and Socolinsky, D. (2003a). Classification using class cover catch digraphs. *Journal of Classification*, 20(1):3–23.

Priebe, C. E., Solka, J. L., Marchette, D. J., and Clark, B. T. (2003b). Class cover catch digraphs for latent class discovery in gene expression monitoring by DNA microarrays. *Computational Statistics & Data Analysis on Visualization*, 43-4:621–632.

Sen, M., Das, S., Roy, A., and West, D. (1989). Interval digraphs: An analogue of interval graphs. *Journal of Graph Theory*, 13:189–202.

Toussaint, G. T. (1980). The relative neighborhood graph of a finite planar set. *Pattern Recognition*, 12(4):261–268.

Tuza, Z. (1994). Inequalities for minimal covering sets in sets in set systems of given rank. *Discrete Applied Mathematics*, 51:187–195.